STUDIES IN HEALTH TECHNOLOGY AND INFORMATICS 235

Informatics for Health: Connected Citizen-Led Wellness and Population Health

Editors: Rebecca Randell
Ronald Cornet
Colin McCowan
Niels Peek
Philip J. Scott

IOS Press

INFORMATICS FOR HEALTH: CONNECTED CITIZEN-LED WELLNESS AND POPULATION HEALTH

Studies in Health Technology and Informatics

This book series was started in 1990 to promote research conducted under the auspices of the EC programmes' Advanced Informatics in Medicine (AIM) and Biomedical and Health Research (BHR) bioengineering branch. A driving aspect of international health informatics is that telecommunication technology, rehabilitative technology, intelligent home technology and many other components are moving together and form one integrated world of information and communication media. The series has been accepted by MEDLINE/PubMed, SciVerse Scopus, EMCare, Book Citation Index – Science and Thomson Reuters' Conference Proceedings Citation Index.

Volume 235

Recently published in this series

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

Informatics for Health: Connected Citizen-Led Wellness and Population Health

Edited by

Rebecca Randell

School of Healthcare, University of Leeds, UK

Ronald Cornet

Department of Medical Informatics, Academic Medical Centre – University of Amsterdam, The Netherlands
Department of Biomedical Engineering, Linköping University, Sweden

Colin McCowan

Robertson Centre for Biostatistics, University of Glasgow, UK

Niels Peek

Health e-Research Centre, Farr Institute of Health Informatics Research, University of Manchester, UK

and

Philip J. Scott

Centre for Healthcare Modelling and Informatics, School of Computing, University of Portsmouth, UK

IOS Press

Amsterdam • Berlin • Washington, DC

ISBN 978-1-61499-752-8 (print)
ISBN 978-1-61499-753-5 (online)
Library of Congress Control Number: 2017938327

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

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

Preface

Founded in 1976, the European Federation for Medical Informatics (EFMI) is the leading organisation in medical informatics in Europe, representing 32 countries. EFMI is a non-profit organisation concerned with the theory and practice of information science and technology within health and health science in a European context. EFMI's objectives are:

- To advance international co-operation and dissemination of information in medical informatics at the European level;
- To promote high standards in the application of medical informatics;
- To promote research and development in medical informatics;
- To encourage high standards in education in medical informatics; and
- To function as the autonomous European Regional Council of IMIA, the International Medical Informatics Association.

In recent years, Europe has seen major investments in research infrastructure for harnessing the potential of routinely collected health data. In the UK, this has led to the establishment in 2013 of The Farr Institute of Health Informatics Research. Funded by the UK's Medical Research Council (MRC) and nine other funders, The Farr Institute is comprised of 21 academic institutions, two MRC units and public bodies such as Public Health England, Public Health Wales and NHS National Services Scotland. Together they form a national research collaboration led by four regional centres: Farr Institute CIPHER, Farr Institute HeRC, Farr Institute London, and Farr Institute Scotland. The Farr Institute aims to be a global leader in health informatics research through scientific discovery and the enhancement of patient and public health. By analysing data from multiple sources and collaborating with the government, public sector, academia and industry, The Farr Institute will unleash the value of vast sources of clinical, biological, population and environmental data for public benefit.

The Medical Informatics Europe (MIE) conference, organised by EFMI, is a key event in the EFMI calendar. The first conference took place in Cambridge, UK in 1978 and now takes place annually. The Farr Institute has also been establishing its own conference series, with the first Farr International Conference taking place in St. Andrews, Scotland, in 2015, followed in 2016 by a second conference held in Swansea, Wales in collaboration with the International Population Data Linkage Network. For 2017, the decision was made to combine the power and established reputational excellence of EFMI with the emerging and innovative research within The Farr Institute community. EFMI, The Farr Institute, and the British Computer Society have worked together to organise Informatics for Health 2017, a joint conference that combines MIE and the Farr International Conference, creating a scientific forum that allows these two communities to share knowledge, insights, and experience, to advance cross-disciplinary thinking, and to stimulate creativity. The conference took place in the city of Manchester in the UK from the 24th to the 26th of April.

The conference received a total of 404 submissions, in the form of both full papers and abstracts, for oral presentation at the conference. The current volume presents the 116 full papers that were presented at the conference, with contributions from 28 dif-

ferent countries. Abstracts that were presented at the conference can be found in Scott, P.J. et al. (2017). Informatics for Health 2017: Advancing both science and practice, J. Innov. Health Inform., 24(1).

We would like to thank all of those who contributed to these proceedings and the success of this important event, in particular the authors who chose to share their work and the reviewers who generously gave their time and expertise. Special thanks go to the Local Organising Committee for their work in organising such a great event.

Rebecca Randell
Ronald Cornet
Colin McCowan
Niels Peek
Philip J. Scott

List of Reviewers

The Scientific Programme Committee would like to thank the following people who reviewed papers submitted to Informatics for Health 2017:

Somayyeh Abedian
Samina Abidi
Emmanuel Kusi Achampong
John Ainsworth
Yasser Alsafadi
Elske Ammenwerth
Suleman Atique
Eustache Muteba Ayumba
Donna Bailey
Gabriella Balestra
Panagiotis Bamidis
Mohamed Ben Said
Sayonara Barbosa
Lejla Begic Fazlic
Mairead Bermingham
Elena Bernad
Andreas Bietenbeck
Bernd Blobel
Olivier Bodenreider
Dominik Brammen
Bernhard Breil
Ruth Breu
Jim Briggs
Ian Brooks
Philipp Bruland
Ann Bygholm
Catherine Chronaki
James Cimino
Carlos Costa
Manfred Criegee-Rieck
Tim Croudace
Elizabeth Cummings
Vasa Curcin
Karen Day
Berry de Bruijn
Sarah Deeny
Nicolette de Keizer
Vincenzo Della Mea
Kerstin Denecke
Thomas Deserno
Peter Diggle
Will Dixon
Nassim Douali
Martin Dugas
Claudio Eccher
Eric Eisenstein
Pia Elberg
Amado Espinosa
Arild Faxvaag
Jesualdo Tomas Fernandez-Breis
Mircea Focsa
Blanca Gallego
Gersende Georg
Mauro Giacomini
Catharine Goddard
Kirstine Rosenbeck Gøeg
Alejandro Rodríguez González
Natalia Grabar
Oscar Gyde
Maria Hagglund
Harald Heinzl
Mira Hercigonja-Szekeres
Marlien Herselman
Jacob Hofdijk
Ursula Hübner
Miron Iancu
Santiago Iriso
Franziska Jahn
Nico Jähne-Raden
Andrew James
Monique Jaspers
Markus Jochem
Owen Johnson
Josette Jones
Martti Juhola
Christian Juhra
James Kariuki
Thomas Karopka
Soudabeh Khodambashi
Petra Knaup
Ann-Kristin Kock-Schoppenhauer
Peter Kokol

Andrzej Kononowicz
Evangelos Kontopantelis
Georgy Kopanitsa
Vassiliki Koufi
Vassilis Koutkias
Dagmar Krefting
Jörn Krückeberg
Matjaz Kukar
Jean-Baptiste Lamy
Nathan Lea
Lenka Lhotska
Siaw-Teng Liaw
Soren Lippert
Nan Liu
Diana Lungeanu
Christian Lüpkes
Martin Lysser
Farah Magrabi
Masaaki Makikawa
Jose Alberto Maldonado
Romaric Marcilly
Louise Marryat
Maurice Mars
Oleg Mayorov
Luca Mazzola
Stephanie Medlock
Frank Meineke
Anne Moen
Ramin Moghaddam
Georgina Moulton
Thomas Müller
Maurice Mulvenna
Hirenkumar Nakawala
Goran Nenadic
Øystein Nytrø
Okure Obot
Andrej Orel
Niels Peek
Sylvia Pelayo
Petra Perner
José Simão de Paula Pinto
Monika Pobiruchin
Mihail Popescu
Fabian Prasser
Hans-Ulrich Prokosch
Vytenis Punys
Usman Qamar
Silvana Quaglini
Rebecca Randell
Chris Robertson
Dave Robertson
Nancy Roderer
Niels Rossing
Etienne Saliez
Antony Sara
Kaija Saranto
Thomas Schabetsberger
Rainer Schmidt
Björn Schreiweis
Philip Scott
Christof Seggewies
Danielle Sent
Walter Sermeus
Abbas Sheikhtaheri
Michael Shifrin
Chris Showell
Murat Sincan
Neil Smalheiser
Tom Sparrow
Martin Staemmler
Sebastian Stäubert
Jürgen Stausberg
Holger Stenzhorn
Lacramioara Stoicu-Tivadar
Vasile Stoicu-Tivadar
Michael Storck
Wu-Chen Su
Frank Sullivan
Stephen Swift
Walter Swoboda
Paul Taylor
Anders Thurin
Shuichi Toyoda
Allan Tucker
Aristides Vagelatos
Christos Vaitsis
Vivian Vimarlund
Hester Ward
Patrick Weber
Martin Wiesner
Klaus-Hendrik Wolf
Maryati Mohd Yusof
Nabil Zary
Richard Zowalla

Contents

2. Health Data Science

This page intentionally left blank

1. Connected and Digital Health

This page intentionally left blank

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-3

Design and Validation of a Platform to Evaluate mHealth Apps

Daniel PHILPOTT [a], Aziz GUERGACHI [a] and Karim KESHAVJEE [a,b,c,1]
[a]*Ryerson University, Toronto, ON*
[b]*InfoClin Inc, Toronto, ON*
[c]*University of Toronto, Toronto, ON*

Abstract. Emerging technologies show great potential in the field of patient care. One such technology is mobile heath applications (mhealth apps), which have exploded in number and variety in recent years, and offer great promise in the ability to collect and monitor patient health data. Despite their apparent success in proliferation and user adoption, these applications struggle to integrate into the primary care system and there is scant information regarding their efficacy to effect patient behavior and consequently health outcomes. In this paper we investigate the potential of a promising clinical evaluation methodology, response adaptive randomized clinical trials, to rapidly and effectively evaluate the efficacy and effectiveness of mhealth apps and to personalize mhealth app selection to individualize patient benefit. Diabetes prevention provides the use case for evaluating the case for and against response-adaptive randomized trials.

Keywords. Mobile Health Applications, Adaptive Clinical Trials, Diabetes Prevention, mhealth apps, evaluation, randomized controlled trials, RCT,

Introduction

Diabetes prevalence in the Canadian population is increasing rapidly and associated costs are also rising, leading to what some experts have termed an 'economic tsunami' [1]. With early detection and strict lifestyle modifications, close to 60% of Type 2 Diabetes cases can be prevented [2, 3].

The Diabetes Prevention Laboratory (DPL) at Ryerson University has developed an algorithm that detects patients at risk of developing diabetes prior to symptom onset through the review of historical electronic medical record (EMR) data [4]. Due to the early nature of this detection, up to eight years ahead of symptom onset, sustained behavioral change is critical to reduce the risk of these patients in developing diabetes. Our group reviewed emerging technologies and methods that show promise in supporting and tracking behavioral change and drawn these together into a proposed health system architecture that includes 1) the primary care provider, 2) a third party telehealth coaching and app management provider, 3) use of mhealth applications used in an integrated manner to effect patient behavioral change and 4) on-going evaluation

[1] Corresponding author, Karim Keshavjee, 111 Gerrard Street East, Rm 301, Toronto, ON, Canada, M5B 1G8. Email: karim@infoclin.ca.

using a response adaptive randomized clinical trial methodology that assists in personalizing interventions to the needs and preferences of individual patients.

The purpose of this study was to review the feasibility and design for efficacy and cost-savings testing of mhealth apps within a proposed workflow (Figure 1). As mhealth apps are rapidly developed and change over time, traditional methods of efficacy testing such as the 2-armed randomized controlled trial (RCT) does not allow the flexibility and rapidity of testing needed to determine the most efficacious apps for modifying patient behavior or that can automate and help lower the costs of coaching.

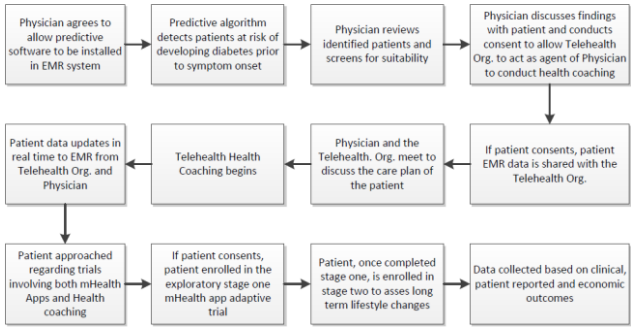


Figure 1. Proposed workflow for mhealth app testing

1. Methods

We conducted interviews with eight key informants from a convenience sample of key stakeholder groups, including researchers, biostatisticians, telehealth program managers and clinicians who could give us feedback on the feasibility and appropriateness of the architecture from their stakeholder perspective. This feedback served to modify the proposed design in an iterative design. In this paper we report on the outcome of these interviews and the collated recommendations with focus on the use and efficacy testing of mhealth apps for the prevention of diabetes.

2. mHealth Apps

2.1. Advantages

Current literature notes the perceived promises of mhealth apps on the ability to positively influence patient lifestyle and behaviors, to increase the data available about the health of patients and to increase adherence to treatment regimens [5, 6]. Indeed several studies comparing patients using various mhealth apps vs. control groups show significant differences in behavioral outcomes when mhealth apps are used [7]. This effect is further enhanced through the concomitant use of healthcare coaching and can enable patients to better manage chronic disease or avoid disease altogether [8].

2.2. Challenges

There are thousands of mhealth apps available for download, but few meet minimum criteria required for integration into the healthcare system. Due to the rapid pace of application development and change, the usual complex and lengthy clinical studies to determine efficacy of a single app simply cannot keep pace with the market. The single app approach also does not account for a varied user base and preferences, does not allow for effective and rapid testing of multiple mhealth apps against each other and does not address data integration with EMRs of healthcare providers.

2.3. Proposed Solution

Mhealth apps should consider the user, system and technological challenges in its design. To do this we propose the use of a response-adaptive randomized (RAR) trial model to allow for the rapid testing of multiple mhealth apps against each other to determine those most efficacious at modifying behavioural change in patients at high risk of developing diabetes. This proposed model was reviewed with our key informants and their feedback allowed for an evolution of this design.

3. Results

Interviewees familiar with the response-adaptive randomized (RAR) trial design had not considered utilizing these types of trials outside of the clinical arena and were intrigued by the idea of applying these to a technology or device.

Four interviewees expressed concern over the complexity of the initial proposed RAR trial design (Figure 1), due to the number of variables involved, which could confound the efficacy data and would both increase cost and complexity. To help address the complexity concerns, two interviewees recommended introducing an 'mhealth app filter' step which would reduce the number of mhealth apps for review. Another interviewee recommended splitting the trial into two separate stages. The first stage would screen the mhealth apps for the most acceptable and user-friendly apps and the second stage to determine the most effective combination of interventions.

The Exploratory Phase (Figure 2) would filter out poorly designed apps based on expert and patient defined criteria. Once a candidate app met expert criteria, it would go through a patient assessment phase using patient reported outcomes as endpoints.

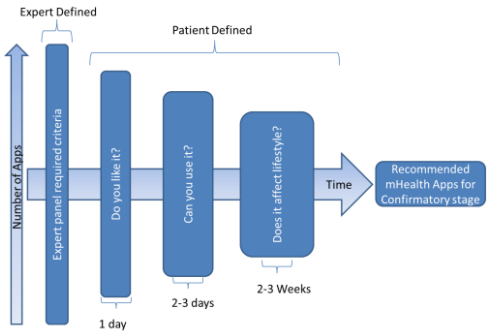


Figure 2. Exploratory Phase

Patients would evaluate apps based on a validated questionnaire, such as the UTAUT2 [9]. Apps that made it through this rigorous gauntlet would be tested for effectiveness in a trial setting.

In the Confirmatory Phase, patients would be randomized into a multi-arm, response adaptive study [10] to test the effectiveness of selected apps against control or in conjunction with other interventions. (Figure 3).

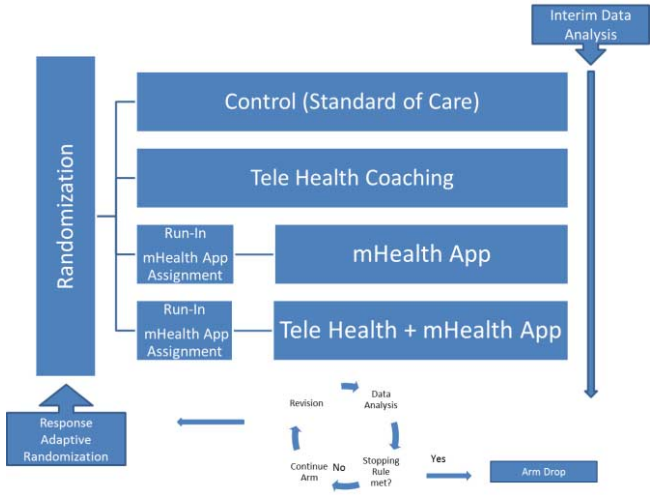


Figure 3. Confirmatory Phase

Effectiveness would be assessed at regular intervals (e.g., 8-16 weeks, depending on how fast the lifestyle modification is expected to show impact based on patient reported, clinical and economic outcomes). For patients where treatment arm was effective, the patient would continue in that arm for an additional period of time. Patients that did not experience an impact would be re-randomized to one of the other arms of the study, favoring arms that had shown greater effectiveness [10]. This design is ethical, rigorous, maximized information and allows researchers to identify the best treatment combination for a patient.

Some key informants mentioned that apps are usually developed for a specific demographic. They recommended use of sub-groups of users to find the most suitable app and not to exclude an app if it is only suitable to one user group (i.e. teenagers). Subgroup user preferences introduced a new level of flexibility into our architecture.

4. Discussion

Key informants provided guidance to enhance the design of response-adaptive randomized trials as part of our proposed diabetes prevention architecture. The initial RAR trial design that was presented was too complex and required updating to ensure it was feasible, not only from a logistical and operational perspective, but also from a review and approval perspective for future agencies and funding bodies. A two stage

design meets the concerns noted well and should still provide the data required to allow for the healthcare architecture to attain proof of concept.

The outcome discussions were very pertinent and have framed the endpoints for the trial design. These outcomes will allow a rich collection of data that will be most useful to answer the questions posed by the healthcare architecture. More thought needs to be given to the constraints, if any, on the control arm of the study(s) as well as how to capture any risk or unanticipated effects the interventions may cause. Other practical considerations that were given, for example the required sample size, will be very useful as the healthcare architecture moves forward. These considerations frame the requirements and scope a little more, although these will be determined in the statistical modeling of a specific RAR trial during the planning phase.

Limitations of the RAR trial include inability to handle changes to apps during the study cycle. Other experimental approaches, such as planned experiments and quality improvement methods [11] may provide better approaches than RAR trials and will also be considered in future studies.

References

[1] Doucet, G., & Beatty, M. (2010). The cost of diabetes in Canada: the economic tsunami. *Canadian Journal of Diabetes*, *34*(1), 27-29.

[2] Diabetes Prevention Program Outcomes Study Research Group, Orchard TJ, Temprosa M, Barrett-Connor E, Fowler SE, Goldberg RB, Mather KJ, Marcovina SM, Montez M, Ratner RE, Saudek CD, Sherif H, Watson KE. Long-term effects of the Diabetes Prevention Program interventions on cardiovascular risk factors: a report from the DPP Outcomes Study. Diabet Med. 2013 Jan;30(1):46-55.

[3] Neamah HH, Sebert Kuhlmann AK, Tabak RG. Effectiveness of Program Modification Strategies of the Diabetes Prevention Program: A Systematic Review. Diabetes Educ. 2016 Apr;42(2):153-65

[4] Mashayekhi M, Prescod F, Shah B, Dong L, Keshavjee K, Guergachi A. Evaluating the performance of the Framingham Diabetes Risk Scoring Model in Canadian electronic medical records. Can J Diabetes. 2015 Apr;39(2):152-6.

[5] Kumar, S., Nilsen, W. J., Abernethy, A., Atienza, A., Patrick, K., Pavel, M., & Hedeker, D. (2013). Mobile health technology evaluation: the mHealth evidence workshop. *American journal of preventive medicine*, *45*(2), 228-236

[6] Hamine, S., Gerth-Guyette, E., Faulx, D., Green, B., Ginsburg, A. (2015) Works citing "Impact of mHealth Chronic Disease Management on Treatment Adherence and Patient Outcomes: A Systematic Review" J Med Internet Res 2015 (Feb 24); 17(2):e52

[7] Naimark, J. S., Madar, Z., & Shahar, D. R. (2015). The Impact of a Web-Based App (eBalance) in Promoting Healthy Lifestyles: Randomized Controlled Trial. *Journal of Medical Internet Research*, *17*(3).

[8] Quinn, C. C., Shardell, M. D., Terrin, M. L., Barr, E. A., Ballew, S. H., & Gruber-Baldini, A. L. (2011). Cluster-randomized trial of a mobile phone personalized behavioral intervention for blood glucose control. *Diabetes care*, DC_110366.

[9] Venkatesh, Viswanath and Thong, James Y.L. and Xu, Xin, Consumer Acceptance and Use of Information Technology: Extending the Unified Theory of Acceptance and Use of Technology (February 9, 2012). MIS Quarterly, Vol. 36, No. 1, pp. 157-178, 2012. Available at SSRN: https://ssrn.com/abstract=2002388

[10] Brown AR, Gajewski BJ, Aaronson LS, Mudaranthakam DP, Hunt SL, Berry SM, Quintana M, Pasnoor M, Dimachkie MM, Jawdat O, Herbelin L, Barohn RJ. A Bayesian comparative effectiveness trial in action: developing a platform for multisite study adaptive randomization. Trials. 2016 Aug 31;17(1):428.

[11] Moen RD, Nolan TW, Provost LP. Quality Improvement Through Planned Experimentation. 2nd ed. New York, NY: McGraw-Hill; 2012.

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-8

Reasoning and Data Representation in a Health and Lifestyle Support System

Sten HANKE[a,1], Karl KREINER[a], Johannes KROPF[a], Marc SCASE[b] and Christian GOSSY[a]
[a] *AIT Austrian Institute of Technology GmbH, Vienna, Austria*
[b] *De Monfort University, Leicester, UK*

Abstract. Case-based reasoning and data interpretation is an artificial intelligence approach that capitalizes on past experience to solve current problems and this can be used as a method for practical intelligent systems. Case-based data reasoning is able to provide decision support for experts and clinicians in health systems as well as lifestyle systems. In this project we were focusing on developing a solution for healthy ageing considering daily activities, nutrition as well as cognitive activities. The data analysis of the reasoner followed state of the art guidelines from clinical practice. Guidelines provide a general framework to guide clinicians, and require consequent background knowledge to become operational, which is precisely the kind of information recorded in practice cases; cases complement guidelines very well and helps to interpret them. It is expected that the interest in case-based reasoning systems in the health.

Keywords. health support system, data reasoning, data representation, health and lifestyle data infrastructure

1. The DOREMI Project

The DOREMI project vision is aimed at developing a systemic solution for healthy ageing, based on a well targeted problem definition and model, able to prolong the functional and cognitive capacity of the older adults by empowering, stimulating and unobtrusively monitoring the daily activities according to well defined active ageing lifestyle protocols. The project joins together the concept of prevention, centred on the older population with a constructive interaction between mind, body and social engagement. The target population had mild cognitive decline and either malnutrition or sedentariness; an integrated monitoring of psychophysiological function combined with promoting cognitive activity, health eating activity, physical activity and social interaction may represent a preventive approach towards further deterioration an onset of new clinical signs.

The DOREMI project enables, through a set of technologies, the principles of independent living and active ageing by empowering older adults to play a greater role in their lifestyle and healthcare decisions and shifting the preventive care implementation and monitoring from institutions to homes with a consequent benefit on both healthcare system savings and improvement in clinical decisions.

[1] Sten Hanke, AIT Austrian Institute of Technology GmbH, Center for Health and Bioresources, Donau-City Straße 1, 1220 Vienna, Austria; E-mail: sten.hanke@ait.ac.at

2. The DOREMI Overall Architecture

From the residence at the pilot site there is a data flow of sensor data from wearables and wireless sensors through the DOREMI system middleware to a raw data database. The subsystems pre-process data (Pre-processing Subsystem), configure the predictive activity recognition tasks (Task Configurator Subsystem), process the daily pre-processed data through the predictive activity recognition components (Activity Recognition Subsystem) and perform exploratory data analysis (EDA Subsystem) [1].

Specifically, social data is produced by processing environmental and wearable sensor information (raw GPS data, PIR data, door contact data and indoor localization) to provide refined high level information (outdoor socialization, indoor socialization, room time aggregation and indoor time aggregation). Similarly, physical activity data is produced based on sensor information collected by the wristband and again further transferred as observations. Dietary data as recorded on the users tablet is first transferred to the MetaDieta database (an external third party datbase of meal data). There, the diet is evaluated and selected results (such as compliance with diet, number of meals per day, daily consumption with fruits, vegetables, calories intake, and daily composition of intake according to bromatologic intake) are further transferred. Gaming data is collected at a specialized game server, aggregated and the results for all users are further transferred as observations. Moreover, games running on the tablet can actively request configuration and difficulty level from the DOREMI reasoning module.

The HOMER DB stores configuration information (e.g. regarding sensors deployment) that is used by the configurator modules to retrieve settings and tuning parameters for the different pilot sites. HOMER[2] is an open and flexible OSGi-based software platform which aims at the integration of various home automation systems and consequential event and situation recognition for smart home (addressing comfort, energy efficiency, etc.) and Ambient Assisted Living (AAL) applications (addressing safety, autonomy, self-confidence, etc.) [2].

The refined data produced by the four activity recognition subsystems is ultimately stored in the KIOLA DB, where it is accessed and exploited by the Reasoner Subsystem to generate aggregated information of interest for lifestyle protocol management. KIOLA is a modular platform for clinical data capture and therapy management, which served as basis for the DOREMI reasoning system as well as the dashboard for supervising physicians. Further, the Reasoner Subsystem provides refined and aggregated information to the end-users and caregivers throughout the physician-side on the DOREMI dashboard.

3. DOREMI Data Reasoning

Figure 1 shows the Reasoner internal dataflow, the high-level database and the dashboard to other DOREMI components and how they are related to the overall system. As seen in previous sections the high-level database receives the data from basically four sources: i) The aggregated activity data, provided by the activity recognition components, supplying information about the physical activity part of the DOREMI protocol as well as parameters of interest for his/her daily assessment. ii) The nutrition data received through a third party database which is managing the food and

[2] http://homer.aaloa.org/

meals taken. iii) The aggregated social data are provided by the unsupervised and supervised activity recognition components and give information on the social interactions of the DOREMI users. The serious games data come from the game environment and are statistics about achievements as well as performance (cognitive performance, speed, memory, attention, flexibility, problem solving etc.).

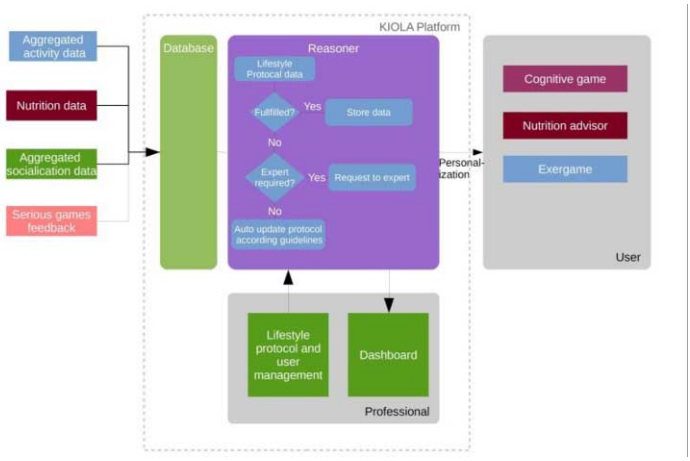


Figure 1. The reasoning module embedded DOREMI environment and other components

The Reasoner is comparing the daily generated subject activity data with the clinical protocol the person should follow based on the pre-sets from the medical experts. Based on this, a general overview as well as some calculated data relations, are presented to the dashboard of the professional. The Reasoner settings are modified in two ways: i) the medical experts decides to change the protocol based on a certain behavior of the user (for example the diet) or ii) the Reasoner adapts the protocol itself (e.g. based on a detected high performance in games it slowly increases the games difficulty).

The Reasoner feedback towards the user is performed in three ways: i) Cognitive games: Achieved goals can be displayed or the person can be informed to change games; ii) Nutrition advisor: Based on changes in the protocol (by medical expert or the Reasoner itself) the person can be advised to change his eating or drinking behavior; iii) Exergames: The person can be motivated and informed about exercises he should do.

The activity recognition modules are placing the generated files in a folder structure. The KIOLA client module would scan the folder structure for new files once a day. In order to guarantee robustness, failed data transmissions would go into a retry-folder or error folder depending on the nature of the error. Once the data has been collected for a given day, the reasoning system starts operating on the data.

4. DOREMI Data Representation

The specialist' dashboard provides specialists with the possibility to review and adjust clinical protocols online. It is designed to be used on both mobile devices and

desktop computers. The dashboard consists of two distinct components: first, the global dashboard provides an overview of all end-users a specialist has access to. Second, the personalized dashboard provides an overview for a specific end-user.

Charts are used to visualize all observations in the area of social, physical game and dietary data. A task module on the dashboard is used to notify specialists when the reasoning system suggests an adoption of the clinical protocol. Here, the specialist can then approve or disapprove the recommendation. Moreover, a specialist can adopt certain parameters of the protocol by himself.

The dashboard has been implemented as web-based application based on the open source framework Django. An apache frontend server is used for communication with browsers. KIOLA itself is a framework layer built on top of Django providing modules for submission of data, dashboard visualizations and the REST API used for communication with other modules. The dashboard itself is based on the open-source CSS framework Bootstrap. The design is based on so-called fluid techniques, meaning the layout adjusts itself based on the device the dashboard is viewed on (e.g. mobile devices).

5. DOREMI System Evaluation

The inclusion criteria for participants of the DOREMI system was people aged between 65 and 80 years old, living alone independently and with mild cognitive impairment and either sedentariness or malnutrition. An iterative user-centered design approach was taken in order to produce an interface that had a high usability for participants. Through a series of four iterative focus groups consisting of participants with the DOREMI demographics the particular interface of the DOREMI system was produced.

The DOREMI system was evaluated in a pilot trial. Fifteen participants were recruited from three residential villages in the UK. The participants were supplied with a tablet computer to present training stimuli, messages from KIOLA and record responses. They also wore a smart bracelet to measure physiological responses and record body movement and had motion sensors installed in their home. There was an initial training phase where participants were instructed on how to use the system. The design of the pilot trial consisted of baseline assessment, a period of two months intervention and then follow-up assessment. During the intervention phase participants were left to interact with the DOREMI system with minimal contact with experimenters. The reason for this approach was to try to produce an autonomous system whereby participants could be left and the reasoning system could decide appropriate stimulus levels. Only in the case where there was equipment failure or for some reason that participants were not interacting with the DOREMI system would experimenters intervene. In order to understand better the experience of participants in the pilot trial a sample of seven residents were interviewed separately before the intervention started and six of those seven were interviewed again after the trial had finished.

During the intervention phase the performance of participants was monitored remotely by specialists who logged into the KIOLA server on a daily basis. Participant interaction with the DOREMI system could be viewed together with progress through the tasks provided. The operation of sensors could be easily assessed together with the outcome of the reasoning system demonstrating how participants moved from one level

of difficulty to another corresponding to performance changes associated with health behaviour change.

Those viewing KIOLA during the trial gave feedback that it was helpful to be able to monitor sensor output as it was obvious when there were issues such as sensor malfunction as the data flow was disrupted. It was easy to spot instances where participants had forgotten to charge portable devices overnight because this produced an interruption of data. Furthermore, if KIOLA showed that participants were interacting less with one part of the DOREMI system compared with another a participant could be asked the reason for that (for example personal preferences). The participant would then be reminded that it was important they engage in all activities. KIOLA was particularly useful in being able to monitor levelling up as game difficulty increased over time. Such a system demonstrated that participants were improving as the intervention proceeded and so a more difficult task was presented.

6. Conclusion

The paper introduced the reasoning system as well as the data representation which is used in the DOREMI project. The data reasoning as well as the data representation in lifestyle and health support systems is an essential component as it provides the information for the experts as well as the user them self which are used for adapting the lifestyle protocol. In the paper we presented a general structure of this two essential components which can easily be used for similar project which providing behaviour change support. The systems provided can be combined with any data mining infrastructure. The tools introduced can be used in a closed loop system to evaluate if the lifestyle has been adapted and also if needed provide the justification for a lifestyle protocol change. The results provided so far show on a general level in a qualitative analysis the usefulness of such systems. Further analysis will be required to see if long-term lifestyle changes in older adults occur through support systems described in the paper.

Acknowledgements

This work has been funded in the framework of the FP7 project ”Decrease of cOgnitive decline, malnutRition and sedEntariness by elderly empowerment in lifestyle Management and social Inclusion” (DOREMI), contract N.611650

References

[1] D. Bacciu, S. Chessa, C. Gallicchio, A. Micheli, E. Ferro, L. Fortunati, F. Palumbo, O. Parodi, F. Vozzi, S. Hanke, J. Kropf and K. Kreiner, Smart Environments and Context-Awareness for Lifestyle Management in a Healthy Active Ageing Framework. *Progress in Artificial Intelligence. Lecture Notes in Computer Science*, vol. 9273, pp. 54-66, 2015.

[2] T. Fuxreiter, C. Mayer, S. Hanke, M. Gira, M. Sili and J. Kropf, A modular platform for event recognition in smart homes, *e-Health Networking Applications and Services* (Healthcom), 2010 12th IEEE International Conference on, pp. 1-6. 2010

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-13

Feasibility of Representing a Danish Microbiology Model Using FHIR

Mie Vestergaard ANDERSEN,[1] Ida Hvass KRISTENSEN, Malene Møller LARSEN, Claus Hougaard PEDERSEN, Kirstine Rosenbeck GØEG, Louise B. PAPE-HAUGAARD
Department of Health Science and Technology, Aalborg University, Denmark

Abstract. Achieving interoperability in health is a challenge and requires standardization. The newly developed HL7 standard: Fast Healthcare Interoperability Resources (FHIR) promises both flexibility and interoperability. This study investigates the feasibility of expressing a Danish microbiology message model content in FHIR to explore whether complex in-use legacy models can be migrated and what challenges this may pose. The Danish microbiology message model (the DMM) is used as a case to illustrate challenges and opportunities accosted with applying the FHIR standard. Mapping of content from DMM to FHIR was done as close as possible to the DMM to minimize migration costs except when the structure of the content did not fit into FHIR. From the DMM a total of 183 elements were mapped to FHIR. 75 (40.9%) elements were modeled as existing FHIR elements and 96 (52.5%) elements were modeled as extensions and 12 (6.6%) elements were deemed unnecessary because of build-in FHIR characteristics. In this study, it was possible to represent the content of a Danish message model using HL7 FHIR.

Keywords. Health Level Seven, Medical Informatics, Standards, FHIR, eHealth

1. Introduction

During the last decades, an on-going digitalization of hospitals and healthcare organizations has led to the modularized implementation of a landscape of heterogeneous Healthcare IT (HIT) systems [1]. As a consequence, the HIT systems have become islands of clinical information that exists in silos [2]. The increasing demand for seamless and coherent treatment of patients across hospitals, practitioners and laboratories require that clinicians have access to the most recent patient information. These information-access requirements have led to the need for integrated HIT systems that are able to share and reuse the information of other HIT systems, i.e. obtain more than technical interoperability [3]. When aiming for interoperable HIT systems, health information standardization is a core part of the solution [4].

In the Danish healthcare sector, a great variety of HIT systems have been developed on an ad hoc basis to address local needs. To overcome heterogeneity and to ensure technical interoperability, national messaging standards (MedCom messages) are used to support communication across the healthcare sector [2, 5]. However, new Danish health strategies have prioritized international standardization [6].

[1] Email: mievestergaard@hotmail.com

Recently, these strategies resulted in Denmark joining HL7 [7]. HL7 has initiated the development of a next generation standards framework; FHIR - Fast Health Interoperability Resources [8]. FHIR has flexible modeling opportunities and supports structural and semantic interoperability. FHIR models can be used across multiple infrastructures and architectures e.g. REST and SOA, and FHIR models are claimed to be highly reusable [9]. For example, if a laboratory message is developed, the underlying models (called resources) can be reused when developing a new laboratory information system, the same underlying models will still be usable and ensure the semantic consistency of the laboratory domain when migrating from a message-based to a SOA-based architecture. An important characteristic of FHIR resources is that they only contain elements that are implemented in 80 % or more of existing HIT systems. Focus is therefore retained on the most common elements, whereas more specific and local elements must be handled through the use of extensions i.e. the 80/20 guideline [10]. Another important aspect of FHIR is the use of profiles. Profiles are developed to customize resources to specific use cases. In FHIR profiles, restrictions can be enforced on resources, and data types can be specified (including terminology binding). In addition, profiles are the extendable parts of the FHIR standard. The extendibility means that elements needed to meet a use case can be inserted in a profile of a relevant resource, even though the element is not specified or foreseen in the original resource.

In the scientific literature, the first FHIR related research has already been published [4-5, 11-13]. However, evidence is still scarce, and to our knowledge, none of the existing studies report in depth about the modeling challenges, associated with applying the FHIR specification for real life use cases.

The aim of this study is to evaluate the feasibility of expressing a Danish MedCom message model in FHIR and analyze areas requiring significantly structural design changes. We have chosen to take our point of departure in one specific Danish message model, the Danish microbiology model (DMM) [14]. The purpose of the DMM is to exchange lab results. This model contains rather complex interrelated findings, which makes it possible to investigate design challenges. Using a simple model could potentially lead to an underestimation in design challenges.

2. Methods

To investigate representation of content of a DMM using FHIR the information content was modeled as close to the DMM as possible, i.e. mapping each attribute in the DMM to a FHIR attribute. However, we identified cases where DMM excerpts were structured in a way that did not fit into FHIRs way of modeling in an obvious way. These were identified as incompatibilities and alternative ways of modelling these were identified.

To evaluate feasibility, each attribute in the DMM was modeled to a corresponding FHIR attribute in a relevant resource. Exceptions were if the characteristics of the FHIR standard made attributes unnecessary, or if a FHIR attribute that accurately covered the content of the Danish attribute could not be found. In the latter case, the FHIR profile was extended with an extra attribute in a relevant resource. In the end of the modeling experiment the number of existing FHIR attributes used, the number of extended attributes, and the number of attributes not relevant in the FHIR framework were calculated. In addition, when a matched attribute contained a predefined FHIR

value set, we evaluated whether this value set covered the content of the Danish value set. The result is a model that represents the content of the DMM using FHIR attributes.

Authors 1-4 were responsible for the modelling of the DMM to a FHIR profile. Authors 1-4 have received training by an expert in the FHIR standard. Authors 5-6 evaluated the modeling, based on their experience in health terminology and models research.

3. Results

In Figure 1, an overview is provided that shows the FHIR resources used to represent the content of the DMM. The model is simplified, so that it does not show all the ten linked observations needed to represent the complex microbiology content. The model shows that eight different resources are utilized. The message header e.g. contains information about date and receiver and sender. The clinical data carried in the message is defined by the data-attribute that links to a diagnostic report. The diagnostic report contains information about the patient, the sample and what kind of analysis that has to be performed. In addition, it has its result defined by the ten linked observations mentioned earlier. These observations hold information about e.g. microscopic and culture findings.

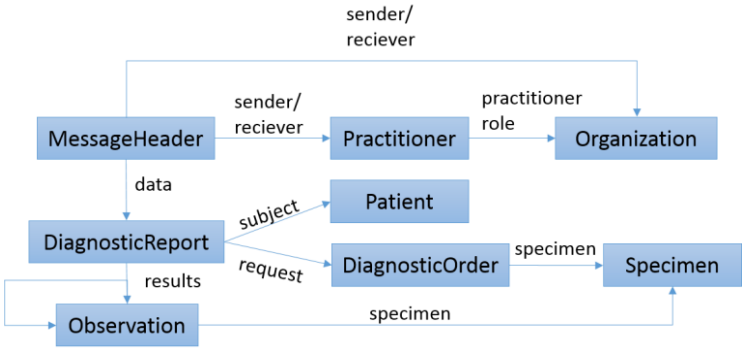


Figure 1. FHIR resources used in the representation of the Danish microbiology model. Arrows are representing references from one resource to another.

The type of information from the DMM controlled whether an element was modeled with an existing FHIR element or whether it was necessary to use an extension. The characteristics of the information that can be modeled directly to an existing FHIR element is information of general character i.e. information which is expected to be used in clinical information systems internationally. The characteristics of the information that needs to be modeled as extensions is information of local or national character. The number of elements from the DMM that is modeled as an existing FHIR element and the number of elements that is modeled as an extension is shown in Table 1.

Table 1. Overview of the distribution of the modeled elements.

Elements modeled as existing FHIR attributes	75 (40.9%)
Elements modeled as extensions	96 (52.5%)
Elements not modeled in FHIR	12 (6.6%)
Total	183 (100%)

Besides from minor design challenges, two major incompatibilities between the DMM and the FHIR model were identified. One was the representation of a table containing the results of a culture analysis, where the antibiotic resistance of bacteria is measured. The other was grouping of microscopic findings. Each of these cases is characterized by linking together a number of findings. Modeling these cases too close to the DMM created an overflow of linked observations with almost no content. Staying more true to the ideas of FHIR modeling allowed us to represent the content much simpler.

4. Discussion

Our results showed that it was possible to represent the content of a Danish microbiology message model using HL7 FHIR with approximately half of the attributes being native for FHIR, and the other half being extensions.

4.1. Extendability

Whenever representing clinical content using standardized models, the extent to which localization is needed should be scrutinized. Extensions means content, and later on patient data, that other systems cannot interpret without access to the local profiles. Given this consideration, the results of this study with half of the attributes represented with extensions, is not very encouraging. However, the percentage of extended clinical content is comparable to other studies representing clinical content using standardized frameworks. For example, Buck et al. (2009) [15] modeled a neonatology electronic patient record using openEHR archetypes and identified 1818 items from the electronic patient record and 1175 items from the paper-based record which were modeled into a total of 132 archetypes. 43.9% (58/132) were reused from available archetypes. Of these 40 archetypes reused as is, 13 were specialized and 12 were extended. 50.8% (67/132) were developed as new archetypes. Even though Buck et al. count per archetype whereas we count per attribute, the additional local content needed is higher or at least comparable to that of our study. While we acknowledge that the complexity of Buck et al.'s study is higher than ours, it shows that getting high re-use percentages is a challenge. To support this, one of our earlier studies also shows that the content overlap between clinical templates in routine use is quite low [16].

Whereas one size fits all would be desirable from a technical viewpoint, it is one of the basic claims of clinical information system research that this is seldom the case e.g. Garde and Knaup (2006) [17]. Consequently, it is very important what kind of content we put in our extensions and what kind of support the standard framework suggests for handling localization of content. In our study, the type of attributes which could not be modeled, were typically content that is only relevant in a local or national context.

4.2. Modeling Approach

In our study, we modeled as close to the original model as possible, because this would allow for the easiest migration. However, the alternative could have been to model as a ”close to FHIR” approach where meaning is modeled instead of attributes. Actively choosing a modeling approach would be an essential part of any infrastructure project where existing models have to be migrated to HL7 FHIR. Our study has the limitation

of not considering the use of standardized terminologies, since standardized terminologies and information models are both a prerequisite in achieving semantic interoperability (e.g. Semantic health report [18]). In our model, local value sets could be replaced with value sets using international classifications or terminologies e.g. SNOMED CT [19]. Using international standards in value sets improve the transfer of meaning, and thus the possibilities of a receiving organization to interpret the transferred patient information correctly.

In conclusion, it was possible to represent the content of a Danish message model using HL7 FHIR. Further investigations are needed to get an overview of implications of adopting HL7 FHIR in a Danish context e.g. getting an overview of design and migration cost given different modeling approaches and decisions about where and how to publish national HL7 FHIR profiles.

References

[1] Smits, M., et al., *A comparison of two Detailed Clinical Model representations: FHIR and CDA.* European Journal for Biomedical Informatics, 2015. **11**(2).

[2] Dalsgaard, E., K. Kjelstrøm, and J. Riis. *A federation of web services for Danish health care*. in *Proceedings of the 7th symposium on Identity and trust on the Internet*. 2008. ACM.

[3] HIMSS. *Definition of interoperability*. 2013; Available from: http://www.himss.org/files/FileDownloads/HIMSS%20Interoperability%20Definition%20FINAL.pdf.

[4] Bender, D. and K. Sartipi. *HL7 FHIR: An Agile and RESTful approach to healthcare information exchange*. in *Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems*. 2013. IEEE.

[5] Pape-Haugaard, L. *Higher level of interoperabtility through an architectural paradigm shift: A study of shared medication record*. in *2011 4th International Conference on Biomedical Engineering and Informatics (BMEI)*. 2011. IEEE.

[6] Sundhed, D., *National strategi for digitalisering af sundhedsvæsenet 2008–2012*. 2008.

[7] HL7. *Denmark joining HL7*. 2013; Available from: http://www.hl7.org/implement/standards/product_brief.cfm?product_id=92.

[8] FHIR. *FHIR Overview*. 2014; Available from: https://www.hl7.org/fhir/overview.html.

[9] FHIR. *HL7 FHIR introduction*. 2014; Available from: http://www.hl7.org/implement/standards/fhir/summary.html.

[10] Grieve, G., #*FHIR and confusion about the 80/20 rule*, in *HealthIntersections*. 2014.

[11] Alterovitz, G., et al., *SMART on FHIR Genomics: facilitating standardized clinico-genomic apps.* Journal of the American Medical Informatics Association, 2015: p. ocv045.

[12] Lamprinakos, G.C., et al. *Using FHIR to develop a healthcare mobile application*. in *Wireless Mobile Communication and Healthcare (Mobihealth), 2014 EAI 4th International Conference on*. 2014. IEEE.

[13] Kasthurirathne, S., et al., *Towards Standardized Patient Data Exchange: Integrating a FHIR Based API for the Open Medical Record System.* Studies in health technology and informatics, 2014. **216**: p. 932-932.

[14] MedCom, *Det nye gode XML mikrobiologisvar*. 2011.

[15] Buck, J., et al., *Towards a comprehensive electronic patient record to support an innovative individual care concept for premature infants using the openEHR approach.* International journal of medical informatics, 2009. **78**(8): p. 521-531.

[16] Gøeg, K.R., et al., *Content analysis of physical examination templates in electronic health records using SNOMED CT.* International journal of medical informatics, 2014. **83**(10): p. 736-749.

[17] Garde, S. and P. Knaup, *Requirements engineering in health care: the example of chemotherapy planning in paediatric oncology.* Requirements Engineering, 2006. **11**(4): p. 265-278.

[18] Society, E.C.D.-G.f.t.I. and Media, *Semantic Interoperability for Better Health and Safer Healthcare: Research and Deployment Roadmap for Europe. Semantic Health Report January 2009*. 2009: EUR-OP.

[19] FHIR. *Using SNOMED CT with FHIR*. 2014; Available from: http://hl7.org/fhir/snomedct.html.

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-18

Establishment of Requirements and Methodology for the Development and Implementation of GreyMatters, a Memory Clinic Information System

Archana TAPURIA[a,1], Matt EVANS[b], Vasa CURCIN[a], Tony AUSTIN[c], Nathan LEA[c], Dipak KALRA[c]
[a]*King's College London (KCL)*
[b]*Berkshire Healthcare NHS Foundation Trust (BHFT)*
[c]*University College London (UCL)*

Abstract. Introduction: The aim of the paper is to establish the requirements and methodology for the development process of GreyMatters, a memory clinic system, outlining the conceptual, practical, technical and ethical challenges, and the experiences of capturing clinical and research oriented data along with the implementation of the system. **Methods:** The methodology for development of the information system involved phases of requirements gathering, modeling and prototype creation, and 'bench testing' the prototype with experts. The standard Institute of Electrical and Electronics Engineers (IEEE) recommended approach for the specifications of software requirements was adopted. An electronic health record (EHR) standard, EN13606 was used, and clinical modelling was done through archetypes and the project complied with data protection and privacy legislation. **Results:** The requirements for GreyMatters were established. Though the initial development was complex, the requirements, methodology and standards adopted made the construction, deployment, adoption and population of a memory clinic and research database feasible. The electronic patient data including the assessment scales provides a rich source of objective data for audits and research and to establish study feasibility and identify potential participants for the clinical trials. **Conclusion:** The establishment of requirements and methodology, addressing issues of data security and confidentiality, future data compatibility and interoperability and medico-legal aspects such as access controls and audit trails, led to a robust and useful system. The evaluation supports that the system is an acceptable tool for clinical, administrative, and research use and forms a useful part of the wider information architecture.

Keywords. memory clinic system, requirements, methodology, EHR standards

1. Introduction

Good memory clinics are multidisciplinary and holistic, integrating health and social care as well as the voluntary sector (www.nao.org.uk) to meet the needs of patients and

[1] Archana Tapuria, Research Fellow, King's College London, SE1 3QD London, UK; E-mail: archana.tapuria@kcl.ac.uk.

their relatives and carers [1]. Standards for memory clinics are specified in the United Kingdom by the Memory Services National Accreditation Programme [2]. At its conception, much time was spent in capturing clinical information, prescribing dementia drugs and monitoring the treatment. It was also recognized that valuable clinical data could be used for service development, research recruitment and primary research purposes. The importance of research has been a key component of the G8 Dementia Summit pledge to find a cure or disease modifying treatment by 2025 (http://dementiachallenge.dh.gov.uk). Thus, there was a need for a system to be developed for the memory clinics of Berkshire Healthcare NHS Foundation Trust (BHFT) to aid the clinical and administrative processes of assessing, diagnosing, managing and treating patients with cognitive disorders and mental health problems and to facilitate recruitment of patients in clinical trials. This memory clinic system was named 'Greymatters'.

2. Methods

2.1 Development of Requirements Standards for EHRs and Systems and Clinical Knowledge Modelling through Archetypes

Requirements were gathered from meetings with clinicians, pharmacists, and administrative staff from BHFT. Google forms builder was used for specifying requirements and a web-based Wiki named JIRA was developed to be able to post any issues, concerns or suggestions so they could be shared between BHFT and University College London (UCL). The software requirements specifications followed the IEEE Standard (http://www.ieee.org), and an EHR standard ISO EN13606 [3] was used. The international standards for EHRs like EN13606 and HL7(accredited by the American National Standards Institute) are extensive and it was not within the scope of the memory clinic system specification to replicate the entire requirements for such systems. Instead they have been used where possible as a background reference and to validate certain of the specific requirements as they arose, against a wider context. Clinical archetypes (http://www.openehr.org) help to fix the hierarchy and representation of the clinical data reducing the variations in the data representation of a particular clinical domain. The clinical, prescribing and research workflow of BHFT was well studied by our clinical team and then modelled to give the design and framework for the Memory Clinic Information system to be built. The clinical data along with their specifications was formally represented by building relevant Clinical Archetypes [4] using the Archetype editor tool, 'Object Dictionary Client' (ODC) [5] developed by UCL.

2.2 Ethical Approval and Clinical Application Development

Ethical approval was not sought as the proposed development work did not directly involve patients and the implementation was a means to supporting existing care. Furthermore, the exposure to patient identifiable information was no more than the clinical team was required to access as part of their daily work as a clinician.

The application takes advantage of a framework built at UCL based on the EN13606 standard for EHR exchange. Clinical model designs created in the ODC are embedded in Java classes (http://www.oracle.com/technetwork/java/index.html) using

relevant classfile metadata. In use, the application runs in an application container JBoss (www.jboss.com) which is installed with an Object-Relational-Mapping (ORM) tool called Hibernate (www.hibernate.org). This provides a rapid means of creating standards-compatible storage for healthcare data. The server could stand alone and accept access requests from any client that can authenticate using Enterprise JavaBeans (http://www.oracle.com/technetwork/java/javaee/ejb/index.html). However, we have created a screen generation framework that behaves rather as an ORM tool does for a database, examining classfile metadata for aggregation and type information to provide screens based on a clinical model expression automatically. This additional facility provides a complete turnkey application development paradigm entirely driven by the original clinical model expression.

2.3 Memory Clinic Archetypes

The following is a list of the main archetypes developed: Demographics, GP details, Alerts/allergies, Consent, Diagnosis, Clinical registers, Cognitive symptoms, Assessment scales, Mental Health Liaison, Referral Data, Medical summary, Medication, Prescribing and dispensing, Research application screen.

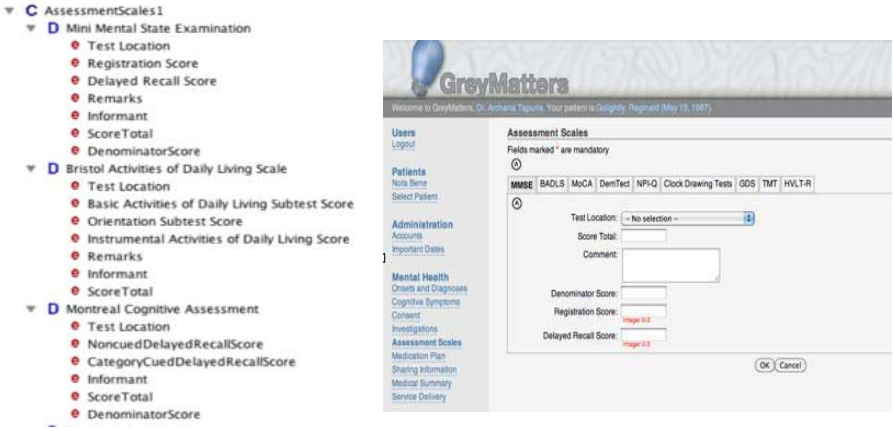


Figure 1. Example of a clinical archetype model and screenshot of the application (Assessment scales)

2.4 System implementation

The application consists of three broad feature sets. The first, for a system administrator, permits “accounts” to be created, “roles”, “users”, and “patients”. An account is a grouping within which users are associated with patients and can see their own care lists. Roles are the named purposes for which users access the care record of a patient. The second feature set is one of care delivery. A set of screens is provided that enable data to be captured in the clinic. Finally, a major component of the system was to offer prompts for repeat prescribing, transfer of prescriptions and dispensing. This area of development proved to be more complex than anticipated and requirements changed as the Trust moved to shared care prescribing. Although built this part of the system was not deployed.

3. Results

The requirements, methodology, technology chosen and standards adopted made the construction, deployment, adoption and population of a memory clinic and research database feasible, along with its research application.

3.1 Deployment and Evaluation of the System

The system was deployed within an NHS managed server environment as a Web application supported by the relational database, PostgreSQL.

To date over 14000 patients are registered on the system. It is currently used by 2 of the 6 locality memory clinics running within the Trust as well as the Research Department and the mental health liaison team for older people in the acute trust. It is primarily used by clinic administrators and secretaries (6 users) who have a key role in recording clinical data, 2 research staff and up to 6 clinicians. In order to integrate the system with the Trust's existing systems and to minimise the burden on administrators, an XML extract is created on a nightly basis of the demographic data of all newly registered patients on the main information system. In an automated process patients are imported into the correct GreyMatters accounts within 24 hours of referral.

An evaluation was undertaken to assess user satisfaction and system usability of GreyMatters in 2014 [6] using the IBM computer usability satisfaction questionnaires [7]. Clinicians and research staff were generally more satisfied than administrators. Overall responses demonstrated mild to moderate satisfaction with the overall system and with individual tasks. Most notably recording of new drug and recording of liaison referral data were considered unsatisfactory. The evaluation shows that the system is an acceptable tool for clinical, administrative, business and research use and forms a useful part of the wider information architecture.

3.2 Research Application

Patient and carer agreement to be contacted about research is recorded in GreyMatters. A previous research database is also being migrated into the GreyMatters database.

From the clinicaltrials.gov research register, it is apparent that the breadth of data items required to identify potential participants for the trials is unlikely to be met by a single system. However, Greymatters dataset captures important data items such as age, sex, diagnosis, assessment scales (e.g. MMSE, BADLS), dementia drug use, antipsychotic use, driving status, and accommodation type, etc. The patient health record including the assessment scales and scores provides a rich source of objective data for audits and research. It helps to establish study feasibility as demographics and 'inclusion and exclusion criteria' can be used in a database search to derive numbers of potential participants. A browser based data mining tool was built on the SQL Server Reporting Platform to pull together data from GreyMatters and other Trust information systems to allow the research department to perform such searches. There are plans to extend the scope to include data from primary care and other secondary care providers.

The latest version of the ODC used to develop the formal information model underpinning GreyMatters is a Web-based application now known as Aruchi, which is recently published [8]. Specific models relating to dementia have been published and are open-source for use (at https://aruchi-helicon.rhcloud.com/pattern/describe?id=134). GreyMatters itself may be licensed in future.

4. Discussion

The development of a system like GreyMatters needs to take into account the existing data collection methods and other information systems that will be used alongside. It was challenging to come to a shared view of the design and requirements of the system given the multidisciplinary nature of the stakeholders, which include Trust members, developers and end users. Significant gaps or errors may not be picked up until systems have been built and are ready for testing or use. The use of the Google forms builder for specifying requirements improved the efficiency and quality of interactions with end users. The benefits of the system have been further enhanced by developing data flows between the different systems. For instance, new patient registrations within the main Trust system are automatically imported into GreyMatters. The browser based recruitment tool for clinical trials added to the benefits of the system. There are plans to roll out GreyMatters further to other locality memory clinics within the Trust.

5. Conclusion

The establishment of requirements and methodology, the importance of the underlying system to address issues of data security and confidentiality, future data compatibility and interoperability and medico-legal aspects such as access controls and audit trails, which led to a robust and useful system. It was beneficial to use a system modelled around standards like IEEE and EN13606 that are based on long established research. This differentiates GreyMatters from simple web based capture forms and this provides the confidence that the system can meet the medico legal challenges of an EHR. The next consideration was to have a flexible approach to capturing clinical data so that it could be reworked to adapt to changing requirements over time. In part this was met by GreyMatters but lack of resources meant that development did not allow the application to exploit certain features to their full potential, for instance the inability to show through the interface successive versions of forms. Its strength is that it provides flexibility to record clinical information. The system has been deemed acceptable by most users although there is dissatisfaction with some aspects and needs further work.

References

[1] Improving services and support for people with dementia. Report by the Comptroller and Auditor General. HC 604 Session 2006-2007, 4 July 2007; National Audit Office.
[2] Memory Services National Accreditation Programme, Standards for Memory Services, Third Edition June 2012, Pub. No. CCQI131Royal College of Psychiatrists.
[3] Kalra D. EN13606 EHR Communication Part 1: Reference Model. CEN TC/251, Brussels. Feb. 2007.
[4] Tapuria A, Kalra D, Kobayashi S. Contribution of Clinical Archetypes, and the Challenges, towards Achieving Semantic Interoperability for EHRs. Healthc Inform Res 2013; 19(4), 286 - 292.
[5] Austin T. The development and comparative evaluation of middleware and database architectures for the implementation of an electronic healthcare record, PhD Thesis. University of London, 2004.
[6] Tapuria A, Evans M, Austin T, Lea N, Kalra D. Development and evaluation of a memory clinic information system. Stud Health Technol Inform 2014; 205:106-10.
[7] Lewis J. IBM Computer Usability Satisfaction Questionnaires: Psychometric Evaluation and Instructions for Use. International Journal of Human-Computer Interaction 1995 v.7 n.1 p.57-78 2007.
[8] Austin T, Sun S, Lea N, Tapuria A, Kalra D. Patterns: a simple but expressive data modelling formalism, Int. J. Knowledge Engineering and Data Mining, Vol. 4, No. 1, 2016.

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-23

Nurses' Perspectives on In-Home Monitoring of Elderlies's Motion Pattern

Joakim KLEMETS[a,1], Jukka MÄÄTTÄLÄ[a], Johan JANSSON[a], and Ismo HAKALA[a]
[a]*University of Jyväskylä*

Abstract. In-home monitoring systems have been proposed to support aging in place and facilitate home care service. Through a qualitative approach the study explores nurses' existing challenges and perspectives of an in-home monitoring system. Results indicate that nurses base care decisions on multiple, and sometimes, unreliable information sources. However, access to information about elderlies' physical motion could support the care planning process by reducing ambiguity and raising attention. Hence, a simple and affordable system that largely relies on nurses to interpret the sensed data could bring additional value.

Keywords. In-home monitoring, homecare, nurse informatics, qualitative research

1. Introduction

The world population is aging rapidly and the number of people above 65 years of age in Europe continues to rise [1]. Cognitive decline, chronicle age-related diseases, as well as limited hearing, vision, and physical ability are all related to aging, creating a significant demand on healthcare services. To enhance healthcare efficiency and support aging in place, a multitude of ambient assisted living and in-home monitoring technologies have been proposed [2-4]. Most of these solutions have been technology driven, focusing on demonstrating technological capabilities. However, the usefulness and effectiveness of these systems from a caregiver's perspective has received less attention [5].

In collaboration with home care services provided by the city of Kokkola, Finland, the goal of the SmartHome4e project[2] is to design, deploy, and evaluate an affordable in-home monitoring system. Utilizing low-cost motion sensors, the system is able to monitor elderlies' in-home physical activity, such as the number of bathroom visits, physical motion, time spent in different rooms (e.g. kitchen, bedroom, living room), and the time spent outside. As an initial step of a design science research process [6], this paper reports on a study that aims to explore nurses' existing care provision challenges and whether elementary information about elderlies' motion patterns could support their work. In particular, through a qualitative approach we investigate the question; *how information about an elderly's physical motion can be utilized by nurses?*

[1] Corresponding author, University of Jyväskylä, Kokkola University Consortium Chydenius, Talonpojankatu 2B, FI-67701 Kokkola, Finland; E-mail: joakim.klemets@chydenius.fi.
[2] Funded by the European Commission and The European Regional Development Fund.

2. Methods

The project adheres to the design science research process model that aims to solve problems through the design of artefacts [7]. Adopting qualitative methods, the study focuses on the two first phases in the design science process, namely; (1) problem identification and motivation, and (2) defining objects of a solution (i.e. what would an artefact accomplish?).

Initially, brainstorming meetings were held with nurses and home care representatives to discuss the idea of an in-home monitoring system. Thereafter, participant observations were carried out to obtain a better understanding of the context of use and nurses' work practices. A researcher shadowed 5 different nurses at work, each occasion lasting about 3 hours. The observations yielded about 15 hours in total.

Based on the gained insight a prototype was developed and used as a vehicle to trigger discussion and reflection of current and future work practices together with nurses [8]. Semi-structured interviews were held with 9 nurses in total. Five of the nurses worked on a home care team that served elderlies living independently in two adjacent apartment buildings. The other four nurses worked on another home care team serving elderlies living in homes distributed throughout the city. During the interviews nurses were presented with the prototype and were asked to reflect on existing work challenges and whether access to information about an elderly's physical motion could support their work. Interviews were transcribed verbatim and analysed together with the observation protocols through a combined inductive-deductive approach [9].

The prototype is designed to be affordable, easy to deploy and maintain, and as little intrusive as possible. Therefore, the system consists of wireless infrared motion sensors (1 per room) and a magnetic sensor attached to the outside door to monitor the elderly's physical motion. The sensed data is sent to a central server for processing and a web application visualises the sensed data to nurses (Figure 1).

The processed data is visualised as daily physical motion information, as well as weekly and monthly views of the physical motion patterns. This provides information such as day/night rhythm, time spent per room, number of bathroom visits, time spent outside of the apartment, etc. The weekly and monthly views provide information about changes in the motion patterns over time.

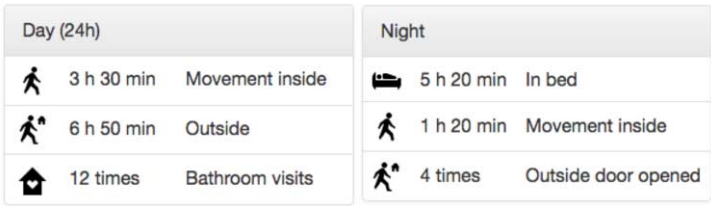


Figure 1. The prototype displaying measured information about an elderly's physical motion.

3. Results

3.1. Challenges with Existing Information Sources

Nurses continuously assess the appropriateness of the care provided and try to detect sudden changes in the elderlies' health. Regardless of any dramatic changes in well-

being, each individual care plan is evaluated twice a year and care needs are estimated together with the elderly. If relevant, whether the elderly should receive care at home or move to a nursing home is also discussed.

Usually, these assessments are based on nurses', relatives', and the elderly's stories about the elderly's well-being. The quest of finding the proper level of care is therefore also complicated by the individuals' motives influencing their story. Through observation and discussion nurses try to develop a thorough picture of the elderly's situation and ability. However, nurse care visits are often short and sometimes infrequent (a few times a month), which makes the task more difficult. A nurse said: 'we struggle to observe that they haven't slept as our visits are so short, at most 30 minutes, and the visits are often in the afternoon, so most of the time we are not aware of that'.

Further, relatives and elderlies tend to want more, or in some cases, less care than what the elderly actually needs. In one example, an elderly couple was withholding information about the husband's frequent falls. The interviewed nurse commented that '...due to fear they didn't want to tell us about the falls. They didn't want to have to receive additional care'. In contrast, nurses also mentioned that they sometimes were confronted with elderlies' asking for more assistance than needed, and even acting incapacitated whenever nurses were around. Nurses stressed that it is important to not provide excessive assistance to maintain the elderly's ability to live independently.

Many of the elderlies also suffer from memory disorders to various degrees, which also make it difficult for nurses to assess the trustworthiness of the information provided. Or as a nurse told: 'we can ask the elderlies, how are you, and then they sometimes answer us almost anything'.

3.2. Raising Attention and Providing Insight

Nurses raised a number of aspects when asked whether access to information about elderlies' physical motion could support their work. Firstly, the information could raise attention to issues of which nurses were unaware, and in this way trigger further investigation. Secondly, the information could help illuminate and clarify already known issues, but where the underlying cause is uncertain. Table 1 summarizes the different aspects brought up during the interviews.

Table 1. Various issues that motion pattern information could raise attention to or further illuminate.

	Issue	**Indicator**
Raise attention	Urinary tract infection	Increased bathroom visits
	Digestion	Increased or prolonged bathroom visits
	Depression	Staying at home, reduced physical motion
	Wandering behaviour	Leaving home irregularly, motion during night
	Pain	Less movement, moving around in bed
	Memory disorder	Developing an irregular motion pattern
Provide insight	Tiredness and fatigue	Irregular sleeping pattern (circadian rhythm)
	Functional ability	Leaving home, moving around
	Pressure ulcer	Stays for long in a similar position
	Hypnotic drug effect	Movement during night
	Nutrition	Time spent in the kitchen, refrigerator door

Regarding the ability to raise awareness, a central issue nurses brought up was the possibility to identify wandering behaviour early on among elderlies suffering from memory disorders. Nurses told several stories of how their clients had been found outside cold, confused, and unable to find their way back, by passers-by or neighbours. A nurse told: 'last week a resident had left his home and was found outside in the morning. Someone had called the ambulance and he was taken to the hospital as he was very cold'. Nurses stressed the importance of implementing interventions early on to prevent wandering, for example, through night visits, medication, or proposing a move to a nursing home.

Nurses also discussed how a trend of reduced physical motion over a period of time could help them to become aware of important issues. Combined with bed relentlessness it could trig the question whether the elderly is in pain. Particularly, if the elderly's pain is already being managed. However, if the elderly is not leaving his/her home as frequently as before, this might also serve as a sign of depression. A nurse told: 'if he spends most of his time in the bedroom and does not leave the apartment that tells me that something needs to be done'. Further, reduced physical motion could also point to a decreased ability to function, which could be supported through more visits or implementing physiotherapy schemes.

Nurses explained that the appropriate care is specifically tailored to each individual's needs and that in the same manner information about elderlies' physical motion needs to be interpreted based on nurses' experiences, the elderly's health history etc. A nurse mentioned a client that was developing a pressure sore due to sitting still for long periods. Due to that elderly's situation the nurse suggested her client to lay down more often rather than, e.g. a physiotherapy intervention.

It became evident that nurses would not rely solely, or mainly, on information captured through sensors when considering care issues. A nurse told: 'we wouldn't rely completely [on physical motion information], I'd interview, ask, and listen as well of course'. However, nurses thought the sensed information could allow them to catch changes in elderlies' well-being earlier. As a nurse explained: 'for those that we visit less regularly it would be beneficial, those that are still in good shape. If there is a sudden change, we could find out earlier. Parkinson's and Alzheimer's disease have these degradation phases'. Further, the information could also support nurses' work through reducing uncertainty about ambiguous issues. Nurses told that if an elderly is looking more fatigued than usual, discovering an irregular circadian rhythm could help explain the issue. Interventions could be implemented such as activating the elderly in the evenings so that he/she does not go to bed too early.

4. Discussion and Conclusion

There is a great variety among the functional capabilities and health conditions of the elderlies served by the home care service teams included in this study. A central issue for nurses is to provide appropriate and timely care according to each individual's needs. The nature of the information sources used in care planning and assessment complicates nurses' work, which often involves detective work and collaboration, requiring problem-solving skills to identify the appropriate level of care.

The results indicate that even straightforward information about elderlies' physical motion gained through sensors can provide valuable insight to the care planning process. Either by raising awareness to a previously unknown issue or by illuminating

already existing, unsolved issues. Hence, comparing the measured information with existing information sources nurses could develop a more complete picture of the elderly's condition and abilities.

However, the results also demonstrate that the measured physical motion needs to be interpreted individually and that a similar motion pattern might not have the same meaning or implications for all elderlies. Rather, nurses' experience, knowledge, and previous interactions with the elderly and their relatives, is essential in interpreting the measured data correctly.

Despite previous efforts to design technically capable in-home monitoring systems [2], the technology readiness for such systems is still considered low [4]. The cost is also a major obstacle for these systems to become widespread [10]. Complex systems, involving a wide range of sensors, further aggravates the technology's maturity and cost. However, our results indicate that nurses can benefit even from a rather simple and affordable system that provides insights to elderlies' motion patterns.

Block et al. [11] highlights the potential of monitoring physical activity for healthcare purposes. However, the authors also conclude that effort is needed to ensure that the measured data is useful to healthcare workers. We argue that a user-centric approach is instrumental in designing cost-effective monitoring solutions that provide valuable, and good enough, information to nurses' work. Gaining an understanding of the context of use through ethnographically inspired methods and involving nurses early in the project has helped us to focus the development of the monitoring system on issues that are important to nurses' work. Further development and evaluation of the monitoring system and its user interface is considered as future work.

References

[1] Eurostat, Population structure and ageing, in, http://ec.europa.eu/eurostat/statistics-explained/index.php/Population_structure_and_ageing.

[2] K.K. Peetoom, M.A. Lexis, M. Joore, C.D. Dirksen, and L.P. De Witte, Literature review on monitoring technologies and their outcomes in independently living elderly people, *Disabil Rehabil Assist Technol* **10** (2015), 271-294.

[3] P. Rashidi and A. Mihailidis, A survey on ambient-assisted living tools for older adults, *IEEE J Biomed Health Inform* **17** (2013), 579-590.

[4] L. Liu, E. Stroulia, I. Nikolaidis, A. Miguel-Cruz, and A. Rios Rincon, Smart homes and home health monitoring technologies for older adults: A systematic review, *Int J Med Inform* **91** (2016), 44-59.

[5] D. Ding, R.A. Cooper, P.F. Pasquina, and L. Fici-Pasquina, Sensor technology for smart homes, *Maturitas* **69** (2011), 131-136.

[6] A.R. Hevner, S.T. March, J. Park, and S. Ram, Design Science in Information Systems Research, *MIS Quarterly* **28** (2004), pp. 75-105.

[7] K. Peffers, T. Tuunanen, C.E. Gengler, M. Rossi, W. Hui, V. Virtanen, and J. Bragge, The Design Science Research Process: A Model for Producing and Presenting Information Systems Research, in: *Proc of DESRIST'06*, 2006.

[8] P. Mogensen and R.H. Trigg, Using Artifacts as Triggers for Participatory Analysis, in: *Proc of PDC'92*, 1992.

[9] A. Tjora, *Kvalitative forskningsmetoder i praksis (Qualitative research in practice)*, Gyldendal Akademisk, 2010.

[10] N. Balta-Ozkan, R. Davidson, M. Bicket, and L. Whitmarsh, Social barriers to the adoption of smart homes, *Energy Policy* **63** (2013), 363-374.

[11] V.A. Block, E. Pitsch, P. Tahir, B.A. Cree, D.D. Allen, J.M. Gelfand, Remote Physical Activity Monitoring in Neurological Disease: A Systematic Review, *PLoS One* **11** (2016).

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-28

Monitoring Activities Related to Medication Adherence in Ambient Assisted Living Environments

Patrice C. ROY[a,1], Samina Raza ABIDI[a, b] and Syed Sibte Raza Abidi[a]
[a] *NICHE Research Group, Faculty of Computer Science, Dalhousie University, Canada*
[b] *Medical Informatics, Faculty of Medicine, Dalhousie University, Canada*

Abstract. A recent trend in healthcare is to motivate patients to self-manage their health conditions in home-based settings. Medication adherence is an important aspect in disease self-management since sub-optimal medication adherence by the patient can lead to serious healthcare costs and discomfort for the patient. In order to alleviate the limitations of self-reported medication adherence, we can use ambient assistive living (AAL) technologies in smart environments. Activity recognition services allow to retrieve self-management information related to medication adherence in a less intrusive way. By remotely monitor compliance with medication adherence, self-management program's interventions can be tailored and adapted based on the observed patient's behaviour. To address this challenge, we present an AAL framework that monitor activities related to medication adherence.

Keywords. Ambient Assisted Living, Activity Recognition, Medication Adherence

1. Introduction

Ambient Assisted Living (AAL) technologies, such as smart environments, are quite suitable for assisting individuals to self-manage their health in home-based setting [1]. The primary function of AAL technologies is to provide adequate and relevant support at the opportune moment – from a healthcare perspective this support can be in terms of monitoring the home-based activities of patients with the intent to remind and guide them about healthcare tasks with respect to their care plan. Ambient services, such as activity recognition, provide the functionality to establish the situational context of an individual and then to provide context-sensitive self-management support to the specific individualized needs of the patient. A recent trend in healthcare is to encourage patients to self-manage their health conditions in home-based settings. Self-management programs guide and motivate patients to achieve self-efficacy in the self-management of their disease through a regime of educational and behavioural modification strategies. In the context of lifetime healthcare, where patients are required to self-manage their condition it is imperative that they follow their prescribed therapy – i.e. adhere to their medication plan by taking their prescribed medications in the right dose and at the right time. Medication adherence is defined as "the extent to which patients follow the instructions they are given for prescribed treatments" [2]. Sub-optimal medication

[1] Corresponding Author: patrice.roy@dal.ca

adherence among patients; particularly those with chronic diseases and receiving long-term therapies, is a problematic issue leading to serious healthcare costs and discomfort for the patient [3]. Although adherence interventions have proven to be effective to a varying degree, almost half of the interventions seem to fail [2]. One of the reasons of sub-optimal outcomes of these interventions is the inability to accurately ascertain whether patients are taking their medications. We argue that one approach to improve medication adherence is to unobtrusively monitor patient's intake of medications at the designated drug intake time and to remind them in case they are non-compliant to their medication regime. We posit that AAL technologies in smart environments can help patients through personalized self-management assistive acts by (a) monitoring a patient's compliance to recommended healthcare activities (such as medication adherence, daily exercise, etc.), and (b) in the event that the patient is forgetful of the recommended activities, reminding the patient to perform the self-management activity (via smartphone local notification, smart home speaker) and recording the completion of the self-management activity. Thus AAL technologies can potentially contribute to the effectiveness and efficacy of self-management programs [4].

In this paper, we present an AAL framework that monitor activities related to medication adherence. In our approach, each activity related to medication adherence is monitored by an activity recognition agent that infers the state of its monitored activity based on relevant observed data collected from connected sensors in a smart environment. To recognize the intended activity of the patient, in our work each activity model is described with a possibilistic network classifier based on possibility theory [5]. The rationale is that we are dealing with uncertain inputs—i.e. the patient may be performing tasks in a non-deterministic manner—hence inferring the activity of a patient requires estimating the possibility of actuation of a series of related actions given the observed inputs. Here, possibilistic network classifiers based on uncertainty theory and using fuzzy sets, possibility theory and non-monotonic reasoning provide a mechanism to handle uncertainty and imprecision of observed data [5]. Activity models use high-level data (contextual information) as inputs, where low-level data-driven components (e.g. action recognition, localization service) provide inputs (contextual data) for the activity recognition task. We have implemented an AAL environment with multiple sensors to monitor a patient's activities, coupled with a series of dedicated activity recognition agents to infer the patient' medication adherence based on the observed activities. The AAL environment was evaluated for 780 medication adherence scenarios with over 80% success in recognizing activities confirming medication adherence.

2. Methods: Activity Recognition for Medication Adherence

We take an AAL approach to monitor a patient's compliance with his/her prescribed medication plan. We have developed an AAL based smart environment comprising (a) sensors to detect the patient's interactions with the environment; and (b) activity recognition agents, developed as possibilistic network classifiers [5], to establish based on the patient's interactions with the environment whether he/she has taken the prescribed medication (i.e. the *Taking Medication* activity). The *taking medication* activity is a composite of a series of sub-activities that need to be completed in a certain temporal order under constraints in order to conclude the taking medication activity.

Given our smart environment, the *taking medication* activity is predicated on the patient taking pills from the pillbox (*Take Pills* sub-activity) and getting drinkable water

in a cup (*Get Water* sub-activity). Both these sub-activities can be carried out in any order but the *Taking Medication* activity duration must be less than 10 minutes, while each sub-activity duration is less than 5 minutes (temporal constraints). In order to take pills from the pillbox in the medicine cabinet, the patients must open the lid (*Open Pillbox* action), turn over the pillbox to remove the pills (*Use Pillbox* action), and close the lid (*Close Pillbox* action). These three actions are carried sequentially (*Open Pillbox* before *Use Pillbox* before *Close Pillbox*) (temporal constraints). In order to get drinkable water to swallow the pills, the patient must retrieve a cup from the cabinet (*Get Cup* sub-activity) and use tap water to fill the cup (*Use Tap Water* sub-activity). The *Get Cup* sub-activity must be carried before the *Use Tap Water* sub-activity (temporal constraints). A scenario where the patient carried out *Taking Medication* and its sub-activities in the kitchen is shown in Figure 1 (path T0 to T4). The proposed approach is designed as a multi-agent AAL system (see Figure 1) where each activity related to medication adherence is monitored by a dedicated *Activity Recognition Agent*. Events collected from the smart environment are sent to the multi-agent system using *Web services*. A *Broker Agent* stores the events (*Events DB*), which are used by *Activity Recognition Agents* in order to monitor activities related to medication adherence. *Self-Management Agents* provide, if needed, assistive notifications via smartphone to improve medication adherence based on the activities' current states and patient's historical behaviour.

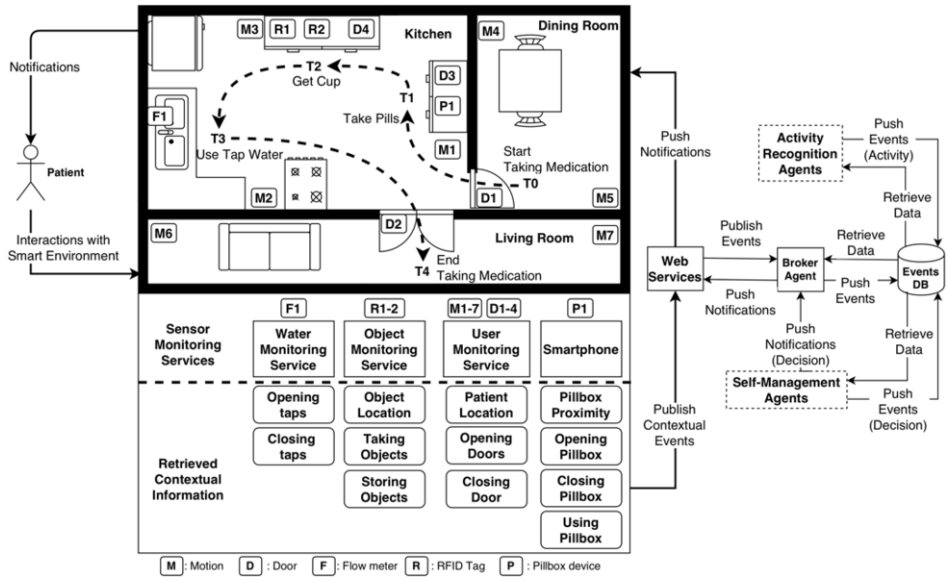


Figure 1. Overview of the medication adherence monitoring system.

2.1. Capturing Patient's Interactions in the Smart Environment

Our smart environment comprises a number of sensors and devices to observe the patient's activities. An observed sensor event occurs when the value of the sensor change (e.g. *Close* to *Open*) or when the sensor's value is published according to a sampling rate (e.g. accelerator value every 0.1 s). We explain below the working of the different sensors. A smart pillbox device (Bluetooth Low Energy) with sensors (accelerometer, reed switch) sends data to a smartphone that is subscribed to events from the pillbox.

Based on those events, the smartphone retrieves contextual information about the patient and the pillbox. The pillbox's accelerometer (orientation) allows to detect if the patient turnovers the pillbox to remove the pills (*Use Pillbox action*). The reed switch sensor on the pillbox's lid allows to observe when the patient opens or close the pillbox (*Open/Close Pillbox action*). Using the smart home's observed events, monitoring services provide contextual information about the patient and the smart environment. Passive infrared sensors (PIR), which detect motions in specific zones, and reed switches on doors allows to detect the patient's location in the smart environment (*Patient Location*). Radio-frequency identification (RFID) tags on objects (e.g. cup or glass) allows to detect the location of object (e.g. kitchen cabinet) and when objects are manipulated by the patient (e.g. *Take Cup*). Flow meter sensors allows to detect when the patient uses water tap in order to get drinkable water (*Use Water Tap*). The retrieved sensor data serves as classifier attributes for the activity recognition agents.

2.2. Activity Recognition Models

Activity recognition (AR) is pursued as a classification task that selects the most plausible class (activity's state) according to the input attributes. We have developed a range of AR models whose goal is to recognize the current activity state (classifier output) based on the abovementioned observed contextual events and attributes (classifier input) during a time interval (*time-window*). A key aspect of our work is that our AR agents are based on possibilistic networks in order to handle certain to uncertain input attributes [5]. This is a practical reality since sensor data can be uncertain/incomplete (e.g. sensor failure) and it is important to account for such uncertainty. In our work, we address this issue by representing an attribute's value with a distribution of certainty degrees on the attribute's plausible values —i.e. complete knowledge (only one certain value) to total ignorance (all values plausible) concerning the attribute value. Each activity model is based on expert knowledge and belief about the activity, where for each classifier attribute (input), the possibilistic conditional distribution of the attribute given the activity's state (output) are a priori defined. Since the observed attributes can be uncertain (distribution instead of a value), the classifier revises the joint distribution in order to take into account this uncertainty and selects the most plausible activity's states, according to the certainty distribution on the possible activity's state values. Once the classification is carried out, a new event with the classification result (selected states or certainty distribution) is created and used for AR.

3. Monitoring Medication Adherence: AAL System in Action

We explain below the working of the AAL system to observe medication adherence, by monitoring the *Taking Medication* activity and its two sub-activities (*Take Pills* and *Get Water*) by activity recognition agents based on dedicated activity recognition models. The recognition of the *Take Pills* sub-activity is based on six attributes: *Pillbox Proximity*, *Open Pillbox*, *Close Pillbox*, *Use Pillbox*, *Patient Location*, and *Temporal Relations*. The patient must carry out three actions (*Open Pillbox*, *Use Pillbox*, and *Close Pillbox*) which must respect the temporal constraints (*Temporal Relations*), while being immediate/near to the pillbox (*Pillbox Proximity*). The recognition of the *Get Water* sub-activity is based on four attributes: *Get cup*, *Use tap water*, *Patient location*, and *Temporal relations*. The patient must carry out two actions (*Get cup*, *Use tap water*), which must respect the

temporal constraints (*Temporal Relations*). The recognition of the *Taking Medication* activity is based on six attributes: *Take pills*, *Get water*, *Temporal relations*, *Patient location*, *Time of the day*, and *Day of the week*. The patient must carry out two sub-activities (*Take pills* and *Get water*), which must respect the temporal constraints (*Temporal relations*), while following the medication schedule (*Time of the day*, *Day of the week*). If the activity is carried out as expected (following the schedule), then the activity is realized in a *normal* way. If the activity occurs outside the expected schedule, then the activity is realized in an *abnormal* way. The AAL system evaluates if the patient behaviour follows the medication schedule and sends a message to the patient to encourage to keep the same behaviour (only *normal* states) or to improve or change his behaviour (missing *normal* states or having *abnormal* states).

4. Results

We implemented the AAL framework and validated its ability to recognize the *Taking Medication* activity using 780 different scenarios of activity realisation (540 for *Taking medication* and 120 for each sub-activity) with six uncertainty levels (from certain observation to complete ignorance about the observation value). For each scenario, we assume that other services generate the uncertain attributes (distributions) that are inputs for the possibilistic network classifiers. The validation results indicate that our AAL system is able to recognize ~79% of the scenarios' activity state. Furthermore, when a scenario's activity state is among the most plausible predicted states, then the system's activity recognition improves significantly to ~98%, which validates our approach of using possibilistic networks to handle uncertain observations.

5. Conclusion

Improving medication adherence allows to alleviate health and economic consequences of non-adherence. We present a unique approach and its implementation for medication adherence that involves monitoring a patient's activities and inferring medication adherence based on their daily activities. A key aspect of our work is the ability to handle the underlying uncertainty when dealing with sensor data and user actions by using possibilistic classifier networks. Validation of our activity recognition models shows that the proposed approach can provide a viable solution for monitoring and helping patients to self-manage their condition, especially helping them with medication adherence.

References

[1] P.C. Roy, A. Bouzouane, S. Giroux, and B. Bouchard, Possibilistic Activity Recognition in Smart Homes for Cognitively Impaired People, *Appl. Artif. Intell.*, **25** (2011), 883–926.
[2] R.B. Haynes, E. Ackloo, N. Sahota, H.P. McDonald, and X. Yao, Interventions for enhancing medication adherence, *Cochrane database Syst. Rev.*, **2** (2008), CD000011.
[3] L. Osterberg and T. Blaschke, Adherence to Medication, *N. Engl. J. Med.*, **353** (2005), 487–497.
[4] T. Patterson *et al.*, Home-Based Self-Management of Dementia: Closing the Loop, in A. Geissbühler *et al.* (Eds) *Inclusive Smart Cities and e-Health*, LNCS 9102, 2015, pp.232–243, Springer, Cham.
[5] S. Benferhat and K. Tabia, Inference in possibilistic network classifiers under uncertain observations, *Ann. Math. Artif. Intell.*, **64** (2012), 269–309.

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-33

Design, Implementation and Operation of a Reading Center Platform for Clinical Studies

Lucien CLIN[a,1], Martin A. LEITRITZ[b], Johannes DIETTER[c], Marek DYNOWSKI[d], Oliver BURGERT[a], Marius UEFFING[c] and Christian THIES[a]
[a] *School of Informatics, Reutlingen University, Germany*
[b] *University Eye-Hospital, University of Tübingen, Germany*
[c] *Institute for Ophthalmic Research University of Tübingen, Germany*
[d] *Now at: Cancer Research UK Manchester Institute, The University of Manchester,UK*

Abstract. Clinical reading centers provide expertise for consistent, centralized analysis of medical data gathered in a distributed context. Accordingly, appropriate software solutions are required for the involved communication and data management processes. In this work, an analysis of general requirements and essential architectural and software design considerations for reading center information systems is provided. The identified patterns have been applied to the implementation of the reading center platform which is currently operated at the Center of Ophthalmology of the University Hospital of Tübingen.

Keywords. Medical Information Systems, Software Platform, Clinical Studies, Telemedicine, Distributed Health Care, Software Engineering

1. Introduction

Reading centers bundle the knowledge of medical experts in a special field to analyse data and provide diagnoses. In the context of clinical studies, reading centers have the potential to ensure consistent high quality by reducing variation and bias via centralized analysis of the data according to fixed SOP [1]. However, this advantage of higher data quality comes at the cost of increased complexity and calls for robust software solutions to address all involved aspects of secure communication, data management, process configuration, visualization, as well as maintenance issues arising over the long-term operation of such a center.

In this work, an analysis of general requirements and ensuing essential architectural and software design considerations is provided, based on the implementation of the reading center platform operated at the Center of Ophthalmology of the University Hospital of Tübingen. The system concept follows the roles and processes defined in IHE Eye Care[2]. But it has been adapted to the needs of the project, e.g. there is no DICOM node available on the data provider side, and data formats were pre-defined by the application field. For that reason, it was considered expedient to

[1] Corresponding author: lucien.clin@reutlingen-university.de

[2] https://www.ihe.net/Eye_Care

develop an open platform that offers flexible modelling of varying distributed reading processes and efficient system interoperability for current standards as well as proprietary components.

2. Method

2.1. Reading Process and System Requirements

The general reading process has been described by Lotz *et al* in [2] as: Data from an examination performed on a *patient* at a *local* site is submitted to a *reading* center, where it is read by a user fulfilling the role of *reader* according to a pre-defined protocol (e.g. study SOP). In some contexts, the data may have to be read independently by two readers ("four eye principle"), with approval by a *senior reader* required in the case of discrepancies. The validated data forms the basis of a *report* returned to the local site.

The use case we are addressing here is an ophthalmology reading center operated at the Center for Ophthalmology of the University Hospital of Tübingen. The data consists of retinal fundus images together with, optionally, visual results or other study-relevant findings (blood pressure, lab results) or background data (questionnaires). Only pseudonymized data is handled within the reading center; all personal identification data remain with the originating local site.

Before entering the reading process, the examination data may be subjected to automated processing. In the present case, for instance, one specific feature of the reading center is the calculation of the artery-to-vein ratio using automated vessel recognition software [3]. For this purpose, the system is configured to submit image-processing jobs to a HPC cluster[3]. From this exemplary process, the following general requirements may be identified for a reading center information system:

1. Integration of data of external origin into a --- potentially pseudonymous --- electronic health record (EHR)
2. Communication with data processing modules (e.g. HPC Cluster) for "further added value"
3. Role- and SOP-based data access, visualization and examination work flow

2.2. System Architecture and Implementation

In the spirit of the "Platform as a paradigm" approach advocated by openEHR [4], in particular the principle that a "platform is a process, not a product", design decisions were primarily grounded in sound software engineering practices, with long-term maintainability in mind. The platform thus follows the service-oriented architecture (SOA) paradigm. In view of the inevitable need for change that arises during the operation use of such a system, a layered design based on established software design patterns has been adopted for loose coupling among individual components and thus minimize implementation effort for modifications. In particular, a metadata approach has been followed for all aspects pertaining to data modelling, from automatic

[3] https://www.bwhpc-c5.de

generation of code for data classes to their flexible mapping onto persistence mechanisms.

The data integration platform (server) has been implemented as a Java web application run on Apache Tomcat[4]; the graphical user interface (GUI) is a rich client written in C++ with the Qt framework[5]. In the following sections, details on the implementation and interplay of the various components are given.

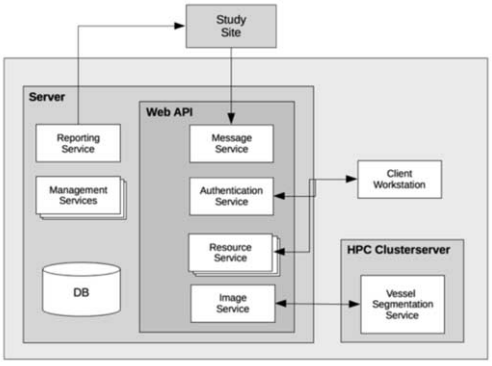


Figure 1. System architecture overview.

2.2.1. Data Exchange

The exchange of data with external sites (i.e. incoming new data to be analyzed, outgoing reports), as well as between the reading center server and GUI-clients, is based on HTTP with XML messages. It is therefore compliant to 'FHIR Messaging' in terms of protocol and serialization format, but ad hoc XML schemata defined for the use case at hand are used instead of adhering to pre-defined FHIR resources. This messaging approach has been chosen instead of a RESTful design because it better suited the event-based processes: external messages with new data initiate the reading process (download of specified images, submission of images to segmentation, notification of readers, etc.); incoming assessments/analyses from readers trigger the comparison for discrepancies or report generation. For secure communication, SSL/TLS is used with mutual server-client authentication based on X.509 certificates.

The Request Mapper Pattern [5] is applied here in order to decouple message formats from internal resource representations and thus allow for their independent evolution. It is implemented using XSLT to transform incoming messages, after which they are unmarshalled into Data Transfer Objects (DTO) [5] using an XML serialization framework[6]. Dealing with a new message format can thus be achieved via mere insertion of an appropriate XSLT script. By this approach interoperability standards, such as the Clinical Document Architecture, are realized as specific configurations for XML serialization.

[4] http://tomcat.apache.org

[5] https://www.qt.io

[6] http://simple.sourceforge.net

2.2.2. Data Modelling and Persistence

The chosen persistence mechanism is a hybrid of relational database (PostgreSQL[7]) and file system storage: A relational schema consisting of the core entities which are to be dealt with in the reading center forms the data model's foundation. All further data is stored as XML-, image- or other files in the file system, references to which are maintained in the relational database. Study-specific data is thus stored in a semi-structured way.

Loose coupling between the server application and the data stores is achieved using a custom persistence framework based on the design presented in [6]. Metadata is here used to generate the code of domain classes and to provide their mapping onto a database.

2.2.3. User Interface

The GUI-client is based on the Application Controller [7] and Model-View-Presenter patterns for modularity and extensibility. The functionality for the aforementioned reader and senior reader modes is implemented using role-specific controller/view components to provide task-specific work flow and visualization: The basic mode guides the reader through the analysis of the patient data (e.g. retinal photographies and other data) according to a study-specific examination catalogue. In the other mode, the senior reader is presented with the patient data and a comparison of the diverging findings with discrepancies highlighted, the successive resolution of which provides an approved report.

3. Results

The presented reading center platform is in operation at the Center for Ophthalmology of the University of Tübingen. After the initial development phase, two small studies with data volumes of 477 and 805 individual examinations (2011 and 3076 images, respectively) were performed on the system by two readers and one senior reader over a period of 7 months; these served as use cases to make final adjustments. At present, the reading center serves to provide examination reports for the ophthalmic data of the "German National Cohort" (NaKo7), a long-scale cohort study. In this context, retinal images of approximately 40.000 patients will be examined over the next years.

4. Discussion

The overall system design and implementation has been validated by successful operation in three studies so far. The applied design principles and technologies (HTTP/XML messaging, O/R-mapping, etc.) are well-established in enterprise application development, and can therefore be directly applied to any clinical information system, irrespective of the domain.

However, the platform is still at a prototypical stage and somewhat tightly coupled to the use cases which guided its implementation. To begin with, many of the processes

[7] https://www.postgresql.org

are hard-coded, in particular in the client GUI. Also, a high level of expert knowledge of the system's inner workings is still required for maintenance operations (user/role management, etc), which need to be performed manually, i.e. at 'system level', for lack of graphical tools. Although, of course, a certain process specificity and administration expertise is inevitable in any information system, we plan to address these issues in a further development iteration. In particular, we intend to extend the metadata approach to dynamic configuration of the client GUI, in particular as concerns study-specific setup of view components and workflow adaptation. In addition, further planned features include a query module (with role-specific access) to access the semi-structured data stored as XML documents; corresponding query forms will here be generated from the modelling metadata. For this purpose, the developed persistence framework is currently being generalized to provide a single API for dealing with further persistence mechanisms, especially the XML database BaseX[8] in addition to relational databases, together with use-case-specific resource representation via configurable serialization mappings. As a general objective, we intend to generalize the core components of this application to provide a basic toolkit for construction of similar platforms.

5. Conclusion

This work analyses essential architectural features and identifies software design patterns for reading center platforms. The proposed design has been implemented in an ophthalmic reading center software system. Since the applied principles and techniques are well-established in enterprise application development, the approach is straightforwardly applicable to other image-based telemedical applications, for instance teledermatology.

References

[1] S.M. Hudson, R. Contreras, M.H. Kanter, S.J. Munz, D.S. Fong, Centralized Reading Center Improves Quality in a Real-World Setting. Ophthalmic Surg Lasers Imaging Retina. 2015;**46**(6):624-629. doi:10.3928/23258160-20150610-05. PMID: 26114842

[2] G. Lotz, T. Peters, E. Zrenner, R.Wilke, A domain model of a clinical reading center - Design and implementation. Conf Proc IEEE EngMed Biol Soc. 2010. 2010:4530-3. doi: 10.1109/IEMBS.2010.5626032.

[3] I.V. Ivanov, M.A. Leitritz, L.A. Norrenberg LA, M. V¨olker, M. Dynowski, M. Ueffing M, J. Dietter, Human Vision-Motivated Algorithm Allows Consistent Retinal Vessel Classification Based on Local Color Contrast for Advancing General Diagnostic Exams. Invest Ophthalmol Vis Sc. 2016; **57**:731-738. doi:10.1167/iovs.15-17831

[4] K. Atalag, T. Beale, R. Chen, T. Gornik, S. Heard, I. McNicoll, openEHR: A semantically-enabled vendor-independent Health Computing Platform, White paper, http://www.openehr.org/resources/white_paper_docs/openEHR_vendor_independent_platform.pdf, last accessed on Nov 07 2016

[5] R. Daigneau, Service Design Patterns: Fundamental Design Solutions for SOAP/WSDL and RESTful Web Services, Addison Wesley, Boston, 2012

[6] S. W. Ambler, The Design of a Robust Persistence Layer for Relational Databases, White paper, http://www.ambysoft.com/downloads/persistenceLayer.pdf, last accessed on Nov 07 2016

[7] M. Fowler, Patterns of Enterprise Application Architecture, Addison Wesley, Boston, 2003

[8] http://basex.org

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-38

Web Validation Service for Ensuring Adherence to the DICOM Standard

Jorge Miguel SILVA [1], Tiago MARQUES GODINHO, David SILVA, Carlos COSTA

DETI/IEETA, University of Aveiro, Portugal

Abstract. The DICOM Standard has been fundamental for ensuring the interoperability of Picture Archive and Communications Systems (PACS). By compiling rigorously to the standard, medical imaging equipment and applications from different vendors can share their data, and create integrated workflows which contributes to better quality healthcare services. However, DICOM is a complex, flexible and very extensive standard. Thus, it is difficult to attest the conformity of data structures produced by DICOM applications resulting in unexpected behaviors, errors and malfunctions. Those situations may be critical for regular PACS operation, resulting in serious losses to the healthcare enterprise. Therefore, it is of paramount importance that application vendors and PACS administrators are confident that their applications follow the standard correctly. In this regard, we propose a method for validating the compliance of PACS application with the DICOM Standard. It can capture the intricate dependency structure of DI- COM modules and data elements using a relatively simple description language. The modular nature of our method allows describing each DICOM module, their attributes, and dependencies on a re-usable basis. As a result, our validator is able to encompass the numerous modules present in DICOM, as well as keep up with the emergence of new ones.

Keywords. DICOM, PACS, Medical Imaging, Validator

1. Introduction

Over the last decades, the use of digital medical imaging systems in healthcare institutions has increased remarkably [1]. Digital medical imaging systems are increasingly becoming central role tools for medical diagnosis and decision support [2]. Research and industry efforts to develop medical imaging equipment, including new acquisition modalities and information systems, are intense and have been grounded by the wide acceptance of the PACS concept [3]. It defines a set of hardware and software technologies that allow standardized data formats and communications between different equipment, applications and information systems [4]. PACS development has been supported by the DICOM Standard. It is the most universal and widespread standard used for the handling, storage, and transmission of digital medical images and related information [5].

The nowadays PACS ecosystem of applications and equipment resorts heavily to the exchange of medical imaging data between them [6]. Therefore, it is crucial that

[1] Corresponding author, Instituto de Engenharia Eletrónica e Informática de Aveiro, Campus Universitário de Santiago 3810-193 Aveiro, Portugal; Email: jorge.miguel.ferreira.silva@ua.pt

PACS components comply rigorously with the DICOM Standard. Since not doing so would compromise the entire operation of the PACS, resulting in potentially serious losses to the medical enterprise [7]. However, due to the complexity of DICOM and the wide variety of supported modalities and information entities, each one with its own set of specifications and dependencies, checking the compliance of an application is not trivial. Therefore, the necessity to conceive a method capable of verifying the compliance of DICOM Objects with the standard comes into sight. Our method checks if the attributes contained in DICOM files are per the standard, namely with the requirements defined in the DICOM Standard modules and templates denominated as Information Object Definition (IOD). It was deployed as a web-application where DICOM files can be uploaded and validated by the community.

2. Methods

DICOM is a Standard that specifies the information content, structure, encoding, and communications protocols for electronic interchange of diagnostic and therapeutic images and image-related information [8]. A central component in the DICOM Standard is the IODs [9], abstract data models used to specify real-world objects. They define which information must be included in each DICOM file, per its object type. Hence, there are IODs defined for different modalities such as the Magnetic Resonance image IOD or the Ultrasound Image IOD. The IOD definition includes a set of modules that contain information about a certain entity, for instance, Patients or Studies. The inclusion of these modules could be mandatory, conditional or optional. The information contained in each module is also defined in DICOM standard Part 3 [9].

Inside the modules, information is convoyed within data elements or attributes. These follow a TLV (Tag Length Value) structure and the standard provides a Data Dictionary that describes all possible attributes. There is also a VR (Value Representation) element that specifies the encoding of each attribute. These are 27 data types in DICOM. It defines the content type, including the characters allowed and prescribed data length. Besides the VR, each data element has also a multiplicity value which defines how many values the element may hold [10].

The wide variety and complexity of DICOM IODs, created the demand for the conception of software capable of automatically verifying attributes and organizations of DI COM files, per the requirements of IODs and Modules defined in the standard.

The dcm4che3 validator[2] is one of the best solutions that uses an undocumented XML file structure to assert IOD validation and verifies if mandatory attributes are de- fined or presented, as shown in this example 2. This structure consists of a root element; the IOD. The IODs children are the Data Elements, which have an associated Tag, VR just like the standard. This validator also supports the definition of a list of acceptable values for the data elements. This allows the definition of attributes such as the Patient's Sex (0010,0040) that must only contain one of" M"," F"," O". It is also capable of sup- porting conditional elements by using If, And, Or conditional clauses, allowing the definition of dependencies such as" Required if Responsible Person is present" in the Responsible Person Role attribute (0010,2298). The dicom3tools/dciodvfy is another solution for checking the DICOM objects

[2] https://github.com/dcm4che/dcm4che

conformance. Like dcm4che, it is capable of checking for inconsistent data inside attributes and between attributes[3].

Despite these features, these validators have two major limitations. First, it cannot resolve conditional elements such as" C - Required if contrast media was used in this image" that is present in many IODs, namely in the MR Image IOD. These are what we call static preconditions, which are not dependent on the IOD itself, but rather on the actual examination procedure. The second limitation is related with the complexity of defining an entire configuration file for each IOD. This problem is severely aggravated by the first problem because it may even be necessary to specify many configurations for the same IOD.

```
<IOD>
<Data Element   keyword=" Attribute Name" tag =" xxxxyyyy" vr =" vv" type =" n"><! ––begin   data element––>
<Value>Accepted   Value 1</ Value>
<Value>Accepted   Value 2</ Value>
</ Data Element> <! –– end data element ––>
</ IOD> <! –– end IOD definition ––>
```

3. Results

Our method tackles the problems mentioned above by enabling the definition of static preconditions that are resolved as inputs from the user. Moreover, it uses an enhanced configuration interface that defines both IODs Modules. By doing so, the validation software becomes able to support the re-usage of each module, and their combination to assemble the whole IOD definition. Lastly, it encompasses an online platform that promotes the sharing of modules and IOD definitions. This greatly reduces the necessity for defining new configurations and therefore simplifies the usage of the application.

As depicted in Figure 1, our methodology can be divided in 3 stages. The initial stage consists of a precondition processor. Preconditions, such as" Required if contrast media was used in this image", are defined in the module configuration file, and are handled by the application as a series of questions that are made to the user during the application's runtime. Furthermore, during this stage the different module configurations are combined into a composite IOD configuration automatically. The output of this stage is a customized IOD configuration file, like the dcm4che's IOD configuration file.

The next stage validates the DICOM files against the customized IOD definitions resultant from the first stage. In this process, we have leveraged the actual dcm4che3 validator, which we have extended to support new features. One may consider the customize IOD configuration resultant from the first stage as a super-set of the dcm4che3 IOD definitions. In this process, the data elements and their dependencies are checked in the source file.

The final stage performs a quality assurance validation. In this stage, some attributes in the DICOM file are checked if they comply with each other. For instance, if the Patient's Age matches the Study acquisition date, and the Patients Birth date. These validations are also defined in each modules configuration file.

[3] http://www.dclunie.com/dicom3tools/dciodvfy.html

Our validation system was deployed in a web platform to provide effortless access to a large user community as possible. This platform will also enable the community to create new, or improve existing module definitions.

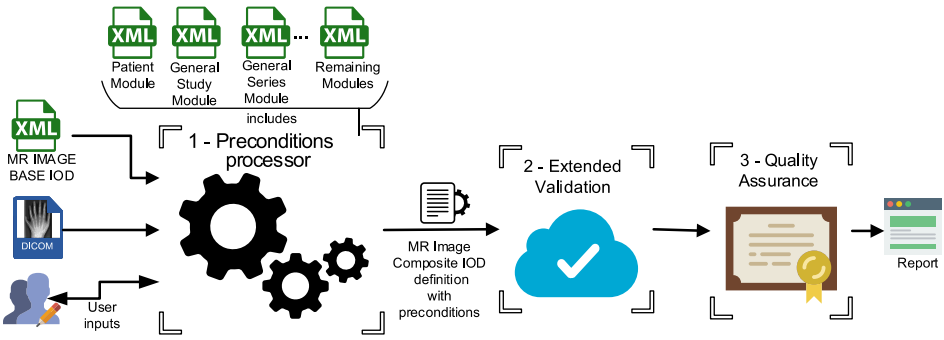


Figure 1. Methodology for the DICOM Validator.

4. Discussion and Conclusion

Our method provides a platform for validating DICOM files. This validator is intended for PACS administrators, developers as well as any other user with interest in validating their DICOM files. Its unique concept of community, and validation-as-a-service promises to greatly ease its usage. The possibility of testing preconditions, and reuse of definition in the DICOM modules level also increases its appeal and usability[4]. Figure 2 shows the result of a DICOM File validation of the Patient, study, series and Image modules. To compare it with other software we provided full access to our application where it is possible to test the validator with any DICOM file. The platform also includes some samples in the application portal, for those who might not have access to real DICOM images.

The implementation of an assurance validator will be useful for checking the congruity of the DICOM file between different attributes of the DICOM file. In other words, the validation method will consider the congruence of inter-attribute information. Finally, it is important to emphasize that this validator has novel features for validating data elements based on module configuration files. By making use of these files and the user's input, the validator creates customized IOD configuration file that defines which data elements and their dependencies are checked in the source file. This functionally increases the range of identifiable violations when compared to others re- ported in the literature.

As future work, we intend to develop more features to integrate in the stage 3, namely, check if different DICOM objects do or not belong to the same patient, and offer a way of fixing incorrect data.

[4] Demo at: http://bioinformatics.ua.pt/dicomvalidator

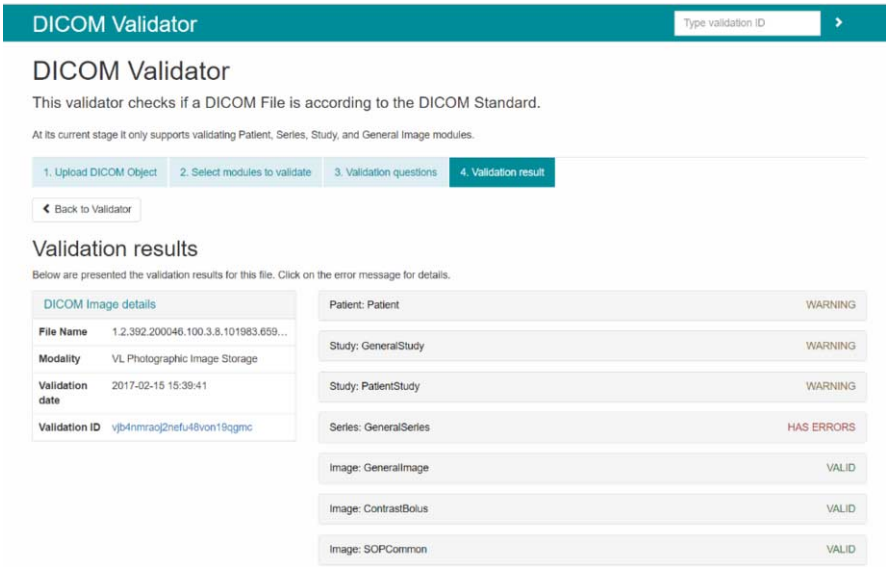


Figure 2. DICOM Validator.

Acknowledgments

Jorge Miguel Ferreira da Silva is funded by the Research grant Project CMUP- ERI / ICT / 0028/2014 - SCREEN-DR, Research Unit 127/94-IEETA. Tiago Marques Godinho is funded by Fundação para a Ciência e Tecnologia (FCT) under grant agreement SFRH/BD/104647/2014.

References

[1] Pedro Matos, Luis A Bastião Silva, Tiago Marques Godinho, and Carlos Costa. A Dynamic Approach to Support Interoperability for Medical Reports Using DICOM SR. Studies in health technology and informatics, 228:461–465, 2016.
[2] E Pinho, T Godinho, F Valente, and C Costa. A Multimodal Search Engine for Medical Imaging Studies. Journal of Digital Imaging, 2016:1–10, 2016.
[3] Frederico Valente, Luis A Bastião Silva, Tiago Marques Godinho, and Carlos Costa. Anatomy of an Extensible Open Source PACS. Journal of Digital Imaging, 29(3):284–296, 2016.
[4] H K Huang. PACS and Imaging Informatics: Basic Principles and Applications. Wiley, 2nd edition, 2010.
[5] Felix Fischer, M. Alper Selver, Sinem Gezer, Ouz Dicle, and Walter Hillen. Systematic parameterization, storage, and representation of volumetric DICOM data. Journal of Medical and Biological Engineering, 35(6):709–723, 2015.
[6] Andrew (Andrew G.) Webb. Introduction to biomedical imaging. Wiley-Interscience, 2003.
[7] Keith J. Dreyer, James H. Thrall, David S. Hirschorn, and Amit Mehta, editors. PACS. Springer-Verlag, New York, 2006.
[8] Oleg S. Pianykh. Digital Imaging and Communications in Medicine (DICOM) - A Practical Introduction and Survival Guide. Springer, 2012.
[9] NEMA. Digital Imaging and Communications in Medicine (DICOM) Part 3: Information Object Definitions.
[10] Herman. Oosterwijk and Paul T. Gihring. DICOM basics. OTech Inc, 2005.

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-43

A Decision Support System for Cardiac Disease Diagnosis Based on Machine Learning Methods

Arash GHAREHBAGHI[a,1], Maria LINDÉN[a] and Ankica BABIC[b,c]
[a] *Department of Innovation, Design and Technology, Mälardalen University, Sweden*
[b] *Department of Information Science and Media Studies, University of Bergen, Norway*
[c] *Department of Biomedical Engineering, Linköping University, Sweden*

Abstract. This paper proposes a decision support system for screening pediatric cardiac disease in primary healthcare centres relying on the heart sound time series analysis. The proposed system employs our processing method which is based on the hidden Markov model for extracting appropriate information from the time series. The binary output resulting from the method is discriminative for the two classes of time series existing in our databank, corresponding to the children with heart disease and the healthy ones. A total 90 children referrals to a university hospital, constituting of 55 healthy and 35 children with congenital heart disease, were enrolled into the study after obtaining the informed consent. Accuracy and sensitivity of the method was estimated to be 86.4% and 85.6%, respectively, showing a superior performance than what a paediatric cardiologist could achieve performing auscultation. The method can be easily implemented using mobile and web technology to develop an easy-to-use tool for paediatric cardiac disease diagnosis.

Keywords. Hidden Markov model, decision support system, heart sound, congenital heart disease screening.

1. Introduction

Rapid progresses in information science initiated a technological leap toward development of the decision support systems by which disease assessment became effectively facilitated. Cardiac disease diagnosis is an important topic that came into this context due to its importance as the statistics show that cardiac disease is still the main factor of human mortality. It has also been reflected on by many researchers that the screening accuracy is not as satisfactory as it could be. The accuracy is low in primary healthcare centers, particularly in children who are sent as a precaution measure to children hospitals or specialized centers for cardiac examination. A great majority of the referred individuals are healthy, but not cleared during the first visit. This comes from the fact that cardiac auscultation is considered as the first screening technique, which is a fairly complicated task; a reliable interpretation of heart sounds needs both the experiences and the expertise. It is well-known that discrimination between normal

[1] Corresponding author, Mälardalens högskola, Box 883, 721 23 Västerås, Sweden; E-mail: arash.ghareh.baghi@mdh.e.

physiological heart sounds from the pathological ones is difficult, especially in pediatric cases, as studies show that as many as 70% of children can have physiological murmurs, while only 0.8% of them are born with congenital heart disease. Consequently, development of a decision support system for improving the screening accuracy can be of a special benefit for the global healthcare system.

Early studies of this topic were initiated in 1990th in which artificial intelligence was employed as the mathematical means of extracting medical information from heart sound [1][2]. Neural networks, as well as the several other statistical techniques, have been well-sounded for the classification purposes [3][4]. Our previous studies have led to innovative methods for processing heart murmurs [5][6][7][8]. We have also developed intelligent method for assessing specific heart diseases [9][10][11]. However, development of a robust system for screening children with mild lesions is still considered as a challenge. One of the related challenges is screening of the abnormalities that may not produce any murmur i.e. bicuspid aortic valve, as most of the existing screening methods are based on the murmur classification.

This paper proposes an original method for developing a decision support system for detecting heart disease in children, even when the abnormality is mild. The main focus of the paper is on the processing method for extracting information from the heart sound time series. We proposed a hybrid method for the classification purpose, based on the encouraging results of our previous studies. The current study tailors the method in a way to include both the first heart sound, and the murmur, to cover a broader patient group. The processing method can be implemented on a portable computer to provide a decision support system for primary healthcare centers.

2. Material and methods

2.1. Data Preparation

Heart sound signals were recorded from the children referrals to the echocardiography lab at the hospital of Children Medical Center, Tehran University of Medical Sciences, Tehran, Iran, using a WelchAllyn Meditron stethoscope in conjunction with a DELL laptop. Each signal contains 10 sec duration of heart sound, recorded at a sampling frequency of 44100 Hz. The informed patient consent was obtained from the legal guardians or from the patients according to the Good Clinical Practice. The study complied with the Declaration of Helsinki and had been approved by the local ethic committee. The patient group was defined as the referrals with Bicuspid Aortic Valve (BAV) and Mitral Regurgitation (MR), against a reference group that constituted of the healthy referrals with or without physiological murmurs, named Innocent Murmur (IM) and No Murmur (NM), respectively. Table 1 lists the patient population

Table 1. Patient population

Heart Condition	Number of Patients	Average Age ± STD (years)
Bicuspid aortic valve (BAV)	20	6.6±1.2
Healthy with innocent murmur (IM)	25	6.7±3.7
Healthy without any murmur (NM)	30	8.6±3.4
Mitral Regurgitation (MR)	15	11.8±4.1

All the referrals were investigated by at least one pediatric cardiologists, and also underwent echocardiography, ECG and chest x-ray as the procedural gold standard of the hospital.

2.2. The Processing Method

The processing method is based on the use of our hybrid method, which is a combination of the Hidden Markov Model (HMM) and support vector machine. The HMM has two states, corresponding to the first heart sound (S1) segment and the systolic period after the S_1 segment ($S_{1\text{-}2}$). A temporal window with fixed length of 50 sec, slides over the signal with an overlapping percentage of 75%, to extract information from the signal. Figure 1 illustrates the state model of the method.

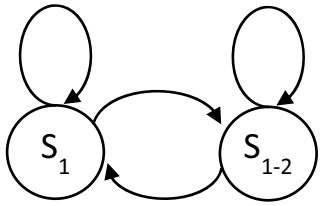


Figure 1. The state model of the method.

Each window is characterized by its spectral contents, using the conventional estimation method, the Priodogram. The Fisher criteria is utilized for finding the most discriminant frequency band, where the spectral contents of the band provide an optimal segregation between the classes [14]. The 8 bands with the highest Fisher value are selected for calculating the spectral energies, which are later mapped to the patterns of numerical symbols, using the Mahalanobis distance. The symbol probabilities are employed as the discriminative features for the classification, which is performed by the support vector machine technique with quadratic kernel. Theoretical foundation for calculating the probability features can be found in [6][8].

2.3. The Statistical Validation

Performance of the method is statistically validated by two different methods, the repeated random sub-sampling (RRSS) and the 5-fold validation method, using the accuracy (I_{ac}), sensitivity (I_{sn}) and specificity (I_{sp}) as the three performance measures:

$$Iac = 100(N_{TP}+N_{TN})/(N_{TP}+N_{TN}+N_{FP}+N_{FN}) \quad (1)$$

$$Isn = 100N_{TP}/(N_{TP}+N_{FN}) \quad (2)$$

$$Isp = 100(N_{TN})/(N_{TN}+N_{FP}) \quad (3)$$

where N_{TP} and N_{TN} are the number of the correctly classified recordings from the abnormal and healthy group, respectively. N_{FP} is the number of the recordings from the healthy group, classified as the abnormal subjects and N_{FN} is the number of the abnormal patient classified as healthy. To apply the RRSS, 50% of the data is randomly selected for training (equally from each class) and the rest for testing the method, after which the performance measures are calculated. This procedure is repeated several times, and the performance measures are calculated accordingly. In a 5-fold validation, each class of data is grouped into 5 divisions, and one division of each class is used for testing and the

rest for training the method. This procedure is repeated 5 times, with one single data being used only once for the testing. At the end, the performance measures are calculated.

3. Results and Discussion

The proposed processing method contains two states, including S_1 and $S_{1\text{-}2}$, for the purpose of screening different conditions i.e. bicuspid aortic valve where the murmur might be missing. The method is trained using the dataset, represented in Table 1, and the average values of the probability features are calculated for each class (see Table 2).

Table 2. The average value of the discriminative features for the 4 classes defined in Table 1

Class	F1	F2	F3	F4	F5	F6
BAV	0.86	0.46	0.19	0.48	0.59	0.21
MR	0.34	0.26	0.29	0.27	0.62	
IM	0.11	0.16	0.22	0.18	0.25	0.69
NM	0.09	0.12	0.15	0.13	0.16	0.17

RRSS was applied to the study data, with 100 iterations, and the performance measures were calculated. Table 3 lists the descriptive statistics of the results.

Table 3. Descriptive statistics of the performance measures

Performance Measure	Average (%)	Median (%)	Standard Deviation
Accuracy	86.4	86.7	3.1
Sensitivity	85.6	85.7	5.3
Specificity	87.0	87.3	3.9

To provide a better representation of the misclassification results, the confusion matrix of the RRSS is listed in Table 4.

Table 4. The average value of the discriminative features for the 4 classes defined in Table 1

	Abnormal (System)	Healthy (System)
Abnormal (Actual)	30	5
Healthy (Actual)	7	48

Previous studies showed that the screening accuracy of a typical pediatric cardiologist is below 80%, relying on the conventional auscultation [10][14][15]. As it could be seen from the Table 3, the accuracy of the method is estimated to be 86.4%, which is significantly higher than the accuracy of a typical pediatric cardiologists. It is also shown that the system is sensitive enough to be employed in primary healthcare centers. In our previous studies, we achieved an accuracy/sensitivity rates of 88%/86% by using the previously proposed method [13]. However, adding the patients with BAV diminished the performance by 6%/10%. Using the proposed hybrid model could sustain the discriminative power at 86.4%/85.6%, thus showing effectiveness of the proposed method to extract relevant information from the signals.

Acknowledgement

The authors gratefully acknowledge the help of Professor Armen Kocharian, Head of

Pediatric Cardiology of the Children's Heart Center of Tehran, Iran and Dr. Amir A. Sepehri for their valuable cooperation in data preparation. This study was supported by the KKS financed research profile Embedded sensor systems for health at Mälardalen University, Sweden and the CAPIS Biomedical Research center.

References

[1] J. Semmlow, M. Akay, and W. Welkowitz, "Noninvasive detection of coronary artery disease using parametric spectral analysis methods,"Engineering in Medicine and Biology Magazine, IEEE, vol. 9, no. 1, pp. 33–36, 1990.

[2] E. Ciaccio, S. Dunn, and M. Akay, "Biosignal pattern recognition and interpretation systems. 3. methods of classification," Engineering in Medicine and Biology Magazine, IEEE, vol. 13, no. 1, pp. 129–135, 1994.

[3] C. G. DeGroff, S. Bhatikar, J. Hertzberg, R. Shandas, L. Valdes-Cruz, and R. L. Mahajan, "Artificial neural network–based method of ng heart murmurs in children," *Circulation*, vol. 103, pp. 2711–2716, 2001.

[4] C. N. Gupta, R. Palaniappan, S. Swaminathan, S. M. Krishnan, Neural network classification of homomorphic segmented heart sounds, Appl. Soft Eng. 7 (2007) 286–297.

[5] A. Sepehri, Amir, A. Kocharian, A. Janani, A. Gharehbaghi, An intelligent phonocardiography for automated screening of pediatric heart diseases, Journal of Medical Systems 40 (2015).

[6] A. Gharehbaghi, P. Ask, A. Babic, A pattern recognition framework for detecting dynamic changes on cyclic time series, Pattern Recognition 48 (2015) 696 – 708.

[7] A. Gharehbaghi, A. Sepehri, A. Kocharian, M. Linden, An intelligent method for discrimination between aortic and pulmonary stenosis using phonocardiogram, in: World Congress on Medical Physics and Biomedi- cal Engineering, June 7-12, 2015, Toronto, Canada, volume 51 of *IFMBE Proceedings*, Springer International Publishing, 2015, pp. 1010–1013.

[8] A. Gharehbaghi, P. Ask, E. Nylander, B. Janerot-Sjoberg, I. Ekman, M. Lindén, A. Babic, A hybrid model for diagnosing sever aortic stenosis in asymp- tomatic patients using phonocardiogram, in: World Congress on Medical Physics and Biomedical Engineering, June 7-12, 2015, Toronto, Canada, volume 51 of *IFMBE Proceedings*, Springer International Publishing, 2015, pp. 1006–1009.

[9] A. Gharehbaghi, P. Ask, M. Lindén, A. Babic, A novel model for screening aortic stenosis using phonocardiogram, in: H. Mindedal, M. Persson (Eds.), 16th Nordic-Baltic Conference on Biomedical Engineering, volume 48 of *IFMBE Proceedings*, Springer International Publishing, 2015, pp. 48–51.

[10] A. Gharehbaghi, T. Dutoit, A. Sepehri, A. Kocharian, M. Linde´n, A novel method for screening children with isolated bicuspid aortic valve, Cardiovascular Engineering and Technology 6 (2015) 546–556.

[11] A. Gharehbaghi, I. Ekman, P. Ask, E. Nylander, B. Janerot-Sjoberg, Assessment of aortic valve stenosis severity using intelligent phonocardiography, International Journal of Cardiology 198 (2015) 58 – 60.

[12] P. N. Belhumeur, J. P. Hespanha and D. J. Kriegman, "Eigenfaces vs. Fisherfaces: recognition using class specific linear projection," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711-720, Jul 1997.

[13] A. Gharehbaghi, M. Borga, B. J. Sjöberg, P. Ask, A novel method for discrimination between innocent and pathological heart murmurs, Medical Engineering and Physics 37 (2015) 674–682.

[14] Watrous RL, Thompson WR, Ackerman SJ. The impact of computer-assisted auscultation on physician referrals of asymptomatic patients with heart murmurs. Clin Cardiol. (2008) Feb;31(2):79-83.

[15] R. Watrous, "Computer-aided auscultation of the heart: From anatomy and physiology to diagnostic decision support," in Engineering in Medicine and Biology Society, 2006. EMBS '06. 28t International Conference of the I EEE, 2006, pp. 140–143.

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-48

Severity Summarization and Just in Time Alert Computation in mHealth Monitoring

Rahul Krishnan PATHINARUPOTHI[a,1], Bithin ALANGOT[b] and Ekanath RANGAN[c]
[a] *Amrita Center for Wireless Networks and Applications, Amrita University, India*
[b] *Center for Cybersecurity Systems and Networking, Amrita University*
[c] *School of Medicine, Amrita Institute of Medical Sciences, Amrita University*

Abstract. Mobile health is fast evolving into a practical solution to remotely monitor high-risk patients and deliver timely intervention in case of emergencies. Building upon our previous work on a fast and power efficient summarization framework for remote health monitoring applications, called RASPRO (Rapid Alerts Summarization for Effective Prognosis), we have developed a real-time criticality detection technique, which ensures meeting physician defined interventional time. We also present the results from initial testing of this technique.

Keywords. mHealth, Critical care alerts

1. Introduction

Remote health monitoring through the use of clinically approved wearable sensors, integrated with the smartphones, is emerging as a promising technological intervention to overcome the lack of affordable access to quality healthcare and timely delivery of critical care. Sensors are now available for monitoring as many as 30 vital cardio-metabolic health indicators, including blood pressure, blood glucose, electrocardiogram, and oxygen saturation to alert any impending critical conditions.

To ensure that critical events are detected and timely alerts generated, we also need to consider the following domain specific challenges too:

- Some of the body parameters change quickly in case of a critical event compared to others. For e.g., an ST level change could occur within few minutes, while a change in blood glucose levels occur over a longer period of time.
- Patient profile: a similar sensor parameter in a younger person could be considered less severe, when compared to an older person.
- We have also observed from experience that the physicians are often reluctant to depend upon a fully automated criticality alerting system, owing to it's wide variability among patients, leading to overwhelming number of alerts.

[1] Rahul Krishnan Pathinarupothi, Amrita Center for Wireless Networks and Applications (AmritaWNA), Amrita School of Engineering, Amritapuri campus, Amrita Vishwa Vidyapeetham (University), India; E-mail: rahulkrishnan@am.amrita.edu

We have developed a rapid summarization and severity detection system that we call RASRPO (Rapid Alert Summarization for Effective Prognosis). It integrates physician feedback with an adaptive severity detection technique, which is able to effectively detect critical events in remote healthcare applications.

Earlier work on identifying trends and discords in large time series data have been discussed in detail by Banaee et al [1], Keogh et al. [2] and Shibuya [5]. Although these techniques have achieved real-time analysis, they are mostly computationally expensive to be run efficiently in smartphones. To our best knowledge, the alert mechanisms already proposed in literature, like the one proposed by Bai et al. [3] are targeted towards generating alarms in an ICU, and could not satisfy all of the above requirements that we set out with. Building on our previous work [4], where we have used a motif-based representation for multi-sensor medical data, we present a novel time-inverted alert and dynamic severity detection technique in this paper.

2. RASPRO Architecture

The patient side architecture begins with sensors attached to human body for measuring and monitoring a variety of physiological parameters. In general, let us consider N vital sensors, $s_1, s_2, \ldots, s_N$, each with a sampling frequency, F. The sampling proceeds continuously for an interval of I time units, following which there may be a gap of Γ time units, and then the sampling resumes for the next interval I, etc. Many such intervals constitute the total observation window Φ. For instance, sampling may occur for I=15 minutes every hour for a day, in which case, Γ = 45 minutes and Φ = 24 hours. The relative durations I, Γ, and Φ are patient and disease specific and are set by the physician. Then a sensor data specific comparator quantizes the digital sequence into one of Q possible severity symbols. For instance, if Q is taken to be five, the levels are labeled A--, A-, A, A+, and A++ with the symbol A indicating normality, and subscripts "-" and "+" indicating sub-normal and above normal levels of increasing severity. The different severity levels are selected from the medical interpretation as well as physician's input based on the patient profile. The severity symbol sequences, all assumed to be of same frequency, are multiplexed at the granularity of one symbol per sensor. Its output is a sequence of timed vectors, with each vector consisting of N values, one from each sensor sampled at that particular instant.

These vectors become the elements of a three dimensional Multi-Sensor Matrix (*MSM*), with $F*I$ columns, and Φ rows, and each element depth equal to N. The MSM can be thought of as consisting of N two dimensional Single Sensor Matrices (*SSM*[1], *SSM*[2], ..., *SSM*[n], ... *SSM*[N]), each of $F*I$ columns and Φ rows. In the next stage of RASPRO, the MSM is used for discovering frequent trends in sensor values that is called consensus abnormality motifs (CAM), which is detailed in our previous work [4]. For the clarity of discussion, we define the following terms below.

2.1. Motifs

Candidate Motif, μ_{CAN}[n] is a temporally ordered sequence of quantized values, $A^*_t, A^*_{t+1}, A^*_{t+2}, \ldots, A^*_{t+L}$ of length L that is selected from SSM [n].

Normal Motif, μ_{NOR}[n] is a candidate motif in which all values represent the normal severity level, which means each and every value is equal to **A**.

Consensus Motif, $\mu_{CON}[n]$ is a candidate motif satisfying the following two conditions: its hamming distance from $\mu_{NOR}[n]$ does *not* exceed a physician prescribed sensor-specific near normality bound, $d_{NOR}[n]$ and, its total hamming distance from all other $\mu_{CAN}[n]$ is the minimum. μ_{CON} represents the observed patient-specific near normal trend.

Consensus Abnormality Motif, $\mu_{CAM}[n]$, is a candidate motif satisfying the following two conditions: its hamming distance from $\mu_{NOR}[n]$ exceeds a physician prescribed sensor-specific near normality bound, $d_{NOR}[n]$ and, its total hamming distance from all other $\mu_{CAN}[n]$ is the minimum.

3. Adaptive Severity Summarization and Alert Computation

3.1. Alert Measure Index

At the end of each observation window Φ_r, for every patient, we define an aggregate alert score, called the ***Alert Measure Index (AMI)*** (see Figure 1). This is calculated as

$$\mathbf{AMI}[\Phi_r] = \Sigma_{i=1}^{N} W[i] * \Sigma_{j=1}^{F*I} \mathbf{num}(\mu_{CAM}[i][j]) * \Theta[j] \quad (1)$$

Wherein, the inner summation takes each severity value in the μ_{CAM} of the i^{th} sensor, converts it into a numerical value (e.g., A± is assigned 1, A++/-- is assigned 2), scales it up by a severity specific factor $\Theta[j]$, and the outer summation scales it up by a sensor specific weightage $W[i]$, both of which are derived from medical domain expertise. We call these two factors W and Θ as severity factors, and the resulting AMI is indicative of the immediacy of patient priority for physician's consultative attention.

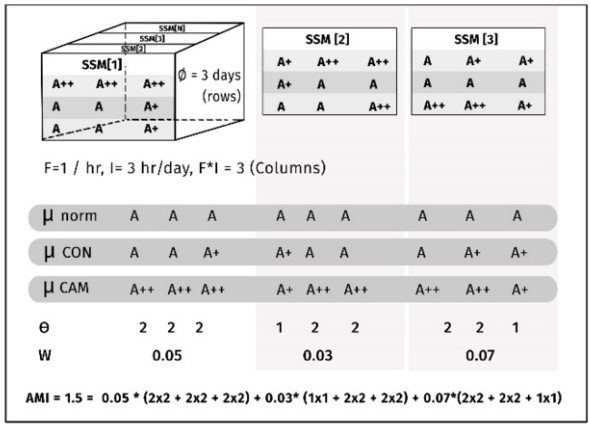


Figure 1. Computation of the Motifs and Alert Measure Index from sensor matrices.

3.2. Interventional Time

The goal of delivering the alerts to the physician is to indicate the upper bound on the time that can elapse before which the physician's intervention is imperative to pull the patient out of danger. In order to capture this, we define the severity factors W and Θ as follows:

$$\Theta[j] = \frac{K1}{\Delta[\alpha]}, \; W[n] = \frac{K1}{\Delta[n]} \quad (2)$$

where, $\Delta[\alpha]$ is the upper bound on the time for intervention for severity level α, $\Delta[n]$ is the upper bound on the time for intervention for sensor n. In (2), constants K_1 and K_2 can be set by the physician considering the context of patient's health condition.

3.3. Adaptive AMI Calculation

AMIs also serve as a feedback mechanism to modulate the frequency of running criticality detection algorithms. A low or high AMI represents a corresponding inversely proportional resultant intervention time as well, referred to $\boldsymbol{T_{INT}}$. Table 1 lists a patient specific data with AMI calculated from only two sensors.

Table 1. AMI and time for intervention are given in this table for a particular patient-disease condition.

BP (mmHg)	Glucose (mg/dL)	AMI	T_{INT} (minutes)
240 / 110	> 420	4 – 5	20
200 / 110	280 – 420	3 – 4	40
180 / 100	200 – 280	2 – 3	80
< 140 / 90	120 – 200	1 – 2	360

We model the frequency of AMI calculation using a linear model as:

$$\boldsymbol{f_{r+1}} = \frac{\boldsymbol{C_1}}{\boldsymbol{T_{INT}(AMI)_r}} + [C_2 * \; (AMI_r - AMI_{r-1})] \quad (4)$$

In (4), C_1 is the intervention time constant in the first term, which ensures that the next AMI is calculated within the current T_{INT}. C_2 is the growth rate constant that modifies the frequency based on rising or falling trend in previous AMI calculations.

4. Implementation and Preliminary Results

We have built an initial implementation of the RASPRO architecture, and carried out preliminary testing of the alerting techniques on anonymized patient data at our 1500-bed super-specialty hospital, namely, the Amrita Institute of Medical Sciences. We have seen very encouraging results during the early trials of the system at the hospital.

Figure 2 shows blood glucose levels measured using continuous glucose monitoring (CGM) interstitial chips from a patient. The continuously collected 24 hour raw values are analyzed for criticality (in terms of AMI) at a fixed frequency and then using the adaptive AMI calculation technique of (4), where C_1=1 and C_2=0.5. The initial frequency was $\boldsymbol{f_1 = 3}$ per hour. The second plot in Figure 2 is the AMI (1 to 5) calculated once in every 20 minutes from the raw sensor values from the CGM chip. We observe that the AMI levels are following expected trend, w.r.t. the actual sensor values. In the third plot, the frequencies vary from a maximum of 3/hour to a minimum of 0.75/hour according to (4). The second plot suggests that fixed frequency is able to pick up all the events in line with the data. However, this is true only if the frequency is

equal to the highest possible one. The third plot is comparable to the second plot, except that AMI calculations are far less and spaced out in time, when AMI is less critical and vice versa, hence providing support for the argument that an adaptive AMI calculation technique performs as good as a fixed one. Similar observations were made in other patient data as well, all though due to space constraints we have omitted from reporting here.

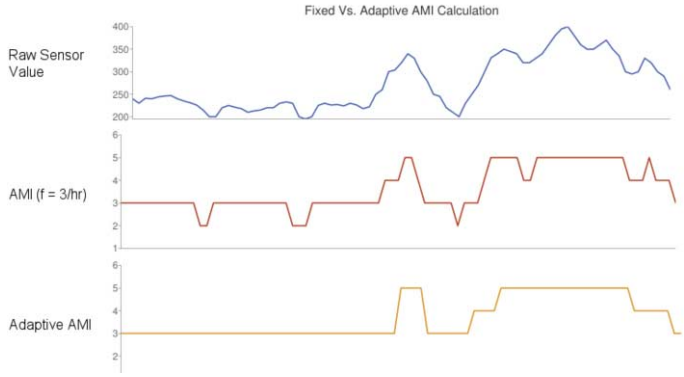


Figure 2. Performance of Adaptive Feedback based Alerts for blood glucose level variations, as compared to fixed frequency alerts.

5. Conclusion

The RASPRO framework combined with a dynamically adapting severity calculation model presented here helps in optimizing criticality detection, balancing the need for timely detection and avoiding redundancy. This will in turn reduce the power and bandwidth requirements for such applications. Results from our initial pilot implementation carried out jointly with practicing physicians are highly encouraging.

References

[1] Banaee H, Ahmed MU, Loutfi A. Data mining for wearable sensors in health monitoring systems: a review of recent trends and challenges. Sensors. 2013 Dec 17;13(12):17472-500.

[2] Keogh E, Lin J, Fu A. Hot sax: Efficiently finding the most unusual time series subsequence. In Fifth IEEE International Conference on Data Mining (ICDM'05) 2005 Nov 27 (pp. 8-pp). IEEE.

[3] Bai Y, Do D, Ding Q, Palacios JA, Shahriari Y, Pelter M, Boyle N, Fidler R, Hu X. Is the Sequence of SuperAlarm Triggers more Predictive than Sequence of the Currently Utilized Patient Monitor Alarms?, IEEE Transactions on Biomedical Engineering (2016) (Issue: 99). IEEE.

[4] Pathinarupothi RK, Rangan E. Discovering vital trends for personalized healthcare delivery. In Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct 2016 Sep 12 (pp. 1106-1109). ACM.

[5] Shibuya N, Nukala BT, Rodriguez AI, Tsay J, Nguyen TQ, Zupancic S, Lie DY. A real-time fall detection system using a wearable gait analysis sensor and a support vector machine (svm) classifier. In Mobile Computing and Ubiquitous Networking (ICMU), 2015 Eighth International Conference on 2015 Jan 20 (pp. 66-67). IEEE.

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-53

Towards Safe and Efficient Child Primary Care – Gaps in the Use of Unique Identifiers in Europe

Grit KÜHNE [a,1], Michael J. RIGBY[b], Azeem MAJEED[a] and Mitch E. BLAIR[b]
[a] *Department of Primary Care and Public Health, Imperial College of Science, Technology, and Medicine, UK*
[b] *Section of Paediatrics Faculty of Medicine, Imperial College of Science, Technology, and Medicine, UK*

Abstract. In order to provide for best possible child health care, timely access to all relevant medical data is of vital importance. The aim of this study is to investigate the use of unique identifiers, a key instrument in this regard, in the countries of Europe. A survey was carried out in all 28 European Member States plus 2 European Economic Area countries in 2015, and refreshed in 2016. In 23 countries unique identifiers are used to link children's health records. Five countries indicated they currently do not link child health records, and two have no such plans. There is variety as regards the type of number and the issuing process.

Keywords. Child health; Intersectoral health care; eHealth and intersectoral documentation, health telematics; Health Data Management and networking

1. Introduction

The UN Convention on the Rights of the Child (to which all European Union (EU) and European Economic Area (EEA) Member States are signatories) defines the highest attainable standard of health care as a fundamental right of every child [1]. The extent to which this requirement is met in practice by national health care systems varies considerably among the countries of Europe, and is the core purpose of the Horizon 2020 funded project Models Of Child Health Appraised (MOCHA), running from 2015 to 2018 [2]. This reported study embedded within MOCHA is focussed on appraising the variations of national health care record linkage systems to identify optimal possibilities, bounded by the ethical and legal concerns regarding the linkage of personal clinical data.

One aspect of the MOCHA study is to examine the role of record keeping in support of primary care, especially electronic child-centric health records, since a precondition for delivery of safe and efficient quality health care is effective and timely access to reliable and inclusive records. Some children, for example neonates, are particularly vulnerable to fragmented clinical information over a comparatively short time, and thus risk receiving sub-optimal care: they are born in one location, supported thereafter by the

[1] Dr. Grit Kühne, Department of Primary Care and Public Health, Imperial College of Science, Technology and Medicine, Reynolds Building, St. Dunstan's Road, London, Hammersmith W6 6RP, United Kingdom; E-mail: g.kuehne@imperial.ac.uk

primary care system and possibly by post-natal midwifery follow-up, may be referred to paediatric specialist services in the event of a health problem, and finally may be taken to emergency services in the event of accidental injury or sudden onset illness. Subsequently, clinical (and social) conditions may change rapidly, necessitating access to timely and complete record systems to ensure that the treating health professional has the full picture of their health history.

However, the reality is often that such vital data are fragmented and locked in different provider-based silo systems. Linked health records can provide the complete picture including possible causes of disease, prior health problems or reactions to previous treatments. Thus they provide information in a far more comprehensive manner than can be expected from a child or parent, or from a single provider record [3].

This is of particular importance for those patient groups who are not themselves in a position to provide this kind of information because of a lack of capacity, be it for reasons of age or illness or because of the intrinsic complexity of multi-provider health care provision. Parents may not always know, or give, a full picture, or may not be present. The importance of record linkage in children's health care has been recognised for over 30 years, e.g. [4-6]

The use of a unique record identifier (URI) is a key instrument in this regard. A URI is a nationally organised number (alpha-numeric or numeric) allocated to each citizen, including children, to link their health records. To be effective for children's record linkage it needs to be given at the time of birth to link all records from that time, though in practice in some countries it is given some time later, compromising safe care.

The URI may be a national citizen ID and used in health, or it may be specific for health only. It is national, in that it follows the child if he/she moves internally. The aim of this study is to investigate the variation in the use and application of time of birth URI systems in the countries of Europe, and possible factors hindering implementation.

2. Methods

A key methodological feature of the MOCHA project is the retention in each study country of a local expert in child health services, who obtains data from local indigenous sources. Questions asked of these Country Agents are passed through internal and external scientific scrutiny to confirm rationale, relevance, and clarity.

Within this frame of the project a survey of unique record identifiers for new-born children was carried out between 14th October and 11th November 2015 (and updated for new data). The questionnaire was designed as a semi structured survey instrument. MOCHA Country Agents were asked to complete the questions on the basis of their expertise, or in cases where this was not possible, to gather data from other sources or national experts on individual questionnaire items. The replies of all participants were analysed using descriptive statistics.

The questionnaire asked whether the country had a specific unique record identifier for children, when this was issued, the structure of the number and what type of records it linked (e.g. was it solely for health, or also a unique citizen identifier for all services). The study also enquired whether the identifier, if it existed, was issued to the parents or to the health system. The questionnaire also asked about whether there were on-going plans or policy debates on strengthening children's record linkage.

3. Results

Replies were initially received from 28 countries. The majority of European countries use a number mechanism to link child health records. Eight of these issue the URI right at birth, the other counties at a later stage. In nine European countries health-specific URI are applied and in 14 more general citizen numbers are used for health record linkage. These 23 countries all use the URI for all records, electronic as well as paper.

Five countries - Austria, Germany, Ireland, Latvia and Slovakia - currently do not link child health records. These five countries differ, however, in the extent to which there are plans for URI implementation. Whereas in Austria, Germany and Ireland there are concrete plans and a set timescale for implementation of a URI including for children, their possible implementation is under debate in Latvia. Slovakia has a number mechanism in place but currently does not use it for record linkage.

Among the countries that link medical records, there is variety as regards the type of number used, the issuing process and details contained as well as the application of URI. Three countries out of 28 reporting differentiated between the application of URI regarding public and private health provision; these are Croatia, Cyprus and Malta.

Figure 1 illustrates the different functionality of URI in different countries, and to whom the number is issued.

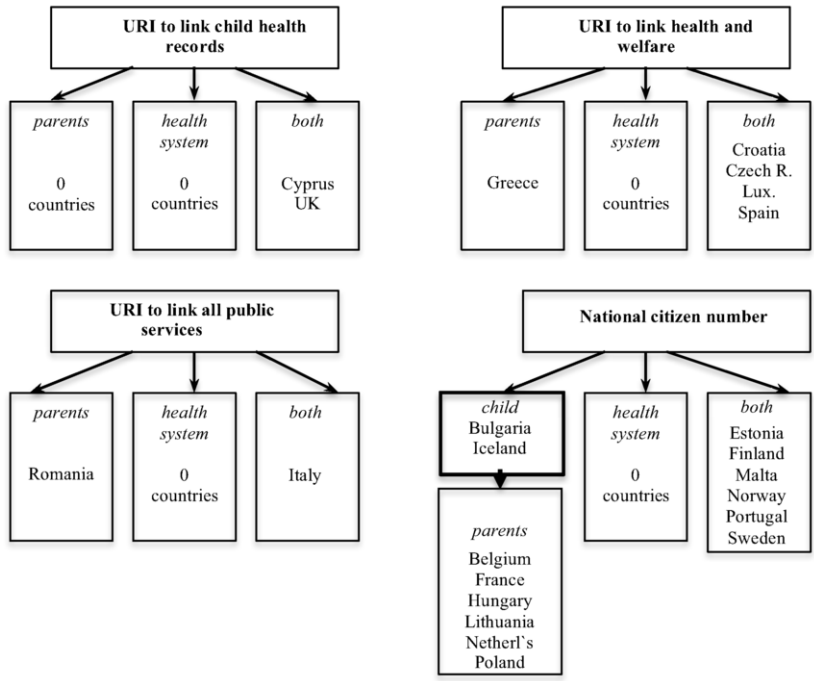


Figure 1. Overview on national functionality and to whom issued

None of the countries issues the identifier only to the health system, which means in all countries with a URI there is awareness, if not full involvement, of parents. Thirteen countries issue the URI to both health system and parents simultaneously, while ten countries issue it to the parents or child, giving them a key role in ensuring linkage.

Whether this is a hindering or supportive factor as regards healthcare access or health record linkage if parents cannot or do not want to provide the URI has not been included in this study.

The assumption that the overall functionality of the number would have an effect on who the identifier is issued to holds true, in that for those countries where the URI is restricted to health, or health and welfare, linkage it is likely to be issued to both the health system and the parents simultaneously (6 out of 7 such countries). By contrast, where it is a national identifier or citizen number, only 6 out of 14 such countries issue it to the health system directly. This adds another step, and thus potential source of delay or error, when seeking to link records swiftly in times of healthcare need.

The Bulgarian and Icelandic policies are exceptional in this respect as the unique identifier is issued to the new-born child directly. Although this at first sight seems unusual as a child will for their first years need adult representation, upon reflection it is an implicit progressive statement that the child is the focus of care, and that the adults are agents for the child, not controllers of the child's health data. Normally parents will act as agents for the child, but this system aids continuity in times of family restructuring.

In Bulgaria, Estonia, Finland, Iceland, Lithuania and the Netherlands the URI is a numerical code issued at birth, in all other countries at a later stage. What was not ascertained in this initial study is how, and how effectively, data about the delivery and early days of the child are linked to ongoing health records if the unique identifier is not available from the start of the child's life. This suggests a potential risk of treatments or preventive measures being based on incomplete data, and requires further study.

Data protection approaches show conflicting attitudes across a spectrum. In some countries data protection is given such gravity that it is a barrier to introducing record linkage, or linkage of clinical items, even if this creates a treatment risk. In some countries a firm middle ground is taken, with linkage only for health, or health and social care, purposes, and an identifier itself containing no data items. At the other end of the spectrum a more liberal approach supports linkage for all public services, use of a general citizen number, and inclusion in the URI of data items (such as date of birth) which stricter countries rate as personal data not to be revealed openly. This study has not considered in more depth the impact and value of these alternative views, and though local values are important further discussion would seem merited.

All these responses provide an insight to the extent that the patterns of record keeping mean that medical decisions in child primary care can be based on timely, accurate and reliable health data. This will aid the MOCHA project in considering how future models of children's healthcare might be optimised.

4. Conclusion

Based on a variety of historic and traditional factors, and the lack of a common vision or standard, there continues to be significant variation in achieving the essential quality enablement of good child health delivery through linkage of early health records. The safety and efficiency of child primary care, as well as trust and convenience for children and parents, is at stake. In every country, this is a national health infrastructure policy determining practice for all citizens including children (and parents). Despite the importance of rapid record continuity, only 8 countries out of 28 enable record linkage from the time of birth, with 3 more planning this. Two countries have no plans for linkage, while 15 have an issuing system implemented some time after birth. These 17 countries

appear prepared to let lack of timely and complete record linkage impede safe care to their children, even in a modern e-health age.

Data protection concerns are deeply rooted and paradoxical. In some countries data protection seems to over-ride safety of health and care. At the other extreme, other countries allow identifiable personal information to be embedded in the identifier. The construct and intrinsic characteristics of the URI also vary considerably between countries. While key criteria are utility and acceptability, the current variation does beg further consideration of whether certain characteristics should prevail. However, and criteria to guide the implementation and properties of URI in child primary care in Europe need to be tailored to the special needs of children and cannot simply be transferred from adult-based systems.

The results obtained within the frame of this first phase of the MOCHA study mark a preliminary insight into the use of unique identifiers in child primary care in Europe. They will serve as a building block for further research on the facilitators and barriers to the development and maintaining effective models of electronic health record support to the delivery of optimal models of child primary care, and will be integrated into the final conclusions of the project in 2018.

Acknowledgements

Acknowledgement is paid to the Country Agents of the MOCHA project as identified on the project website http://www.childhealthservicemodels.eu/partners/). Their contributions ensured that the findings of this study are based on detailed and local indigenous knowledge.

The Models of Child Health Appraised (MOCHA) project is funded by the European Commission through the Horizon 2020 Framework under the grant agreement number: 634201. The sole responsibility for the content of this paper lies with the authors. It does not necessarily reflect the opinion of the European Union, or of the full project. The European Commission is not responsible for any use that may be made of the information contained therein.

References

[1] United Nations, *Convention on the Rights of the Child* (1990), URL https://treaties.un.org/doc/Publication/MTDSG/Volume%20I/Chapter%20IV/IV-11.en.pdf, access date: 06 March 2016

[2] MOCHA Project www.childhealthservicemodels.eu, access date: 07 March 2016

[3] M.J. Rigby (2004). Information as the patient's advocate. In M.J. Rigby (Ed.), *Vision and Value in Health Information* (pp. 57-67), Radcliff Medical Press Ltd., Oxon, 2004.

[4] N.E. Simpson, I.J. Alleslev LJ. Association of children's diseases in families from record linkage data; Can J Genet Cytol. 1972 Dec;15(4):789-800

[5] J.E. Oliver. Successive generations of child maltreatment: social and medical disorders in the parents; Br J Psychiatry. 1985 Nov;147:484-90

[6] C. Sellar, J.A. Ferguson, M.J. Goldacre. Occurrence and repetition of hospital admissions for accidents in preschool children; BMJ. 1991 Jan 5;302(6767):16-9.

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-58

Why Are Children's Interests Invisible in European National E-Health Strategies?

Michael J. RIGBY[a,1], Grit KÜHNE [b], Azeem MAJEED[b] and Mitch E. BLAIR[a]
[a] *Section of Paediatrics Faculty of Medicine, Imperial College of Science, Technology, and Medicine, UK*
[b] *Department of Primary Care and Public Health, Imperial College of Science, Technology, and Medicine, UK*

Abstract. Harnessing the power of IT solutions in child primary care requires strategic thought at national level, and good health care delivery needs this support. The aim of this study was to investigate whether children's needs are considered in national e-health strategies in Europe. In 2016, a survey was carried out in all 28 European Member States plus 2 European Economic Area countries. Sixteen countries fail to mention children's needs at all. Only eleven of 27 countries mention children and adolescents in their national e-health strategy documents ranging from mere data protection concerns to comprehensive IT approaches for the improvement of child primary care.

Keywords. e-health strategies, Child primary care, Europe, informatics, policy

1. Introduction

In May 2012, the Estonian President Toomas Hendrik Ilves, Chair of the independent high-level European Commission's e-Health Task Force said: "*We know that in healthcare we lag at least 10 years behind virtually every other area in the implementation of IT solutions. We know from a wide range of other services that information technology applications can radically revolutionise and improve the way we do things"* [1].

In child health the issues are even more important for a number of reasons. Infants and young children cannot speak for themselves, give their own history, or supply past health information, thus an up-to-date health record is therefore vital. This is even more essential where parents may not know the full details (such as clinical factors at birth), are themselves stressed or confused, or in those cases where for whatever reason the parents neglect their child's best interests. In these cases, as Rigby has previously indicated, the record acts as the child's advocate [2]. Electronic health records give a modern effective way of ensuring that accurate, timely data is available at the point of clinical contact, or in ensuring that all preventive services have reached a child. However, not only is the pattern of implementation of e-health very varied across Europe, but in many cases systems are designed and implemented to a generic model

[1] Prof. Michael J. Rigby, Section of Paediatrics Faculty of Medicine, Imperial College of Science, Technology and Medicine, Reynolds Building, St. Dunstan's Road, London, Hammersmith W6 6RP, United Kingdom; E-mail: m.rigby@imperial.ac.uk

based on adults' needs, without awareness of the particular issues of child e-health records – ranging from the need to link records even before the child has a formal name or civil citizen status, through to the child-specific data items which are important to record such as accurate preventive care (screening and immunisation), growth analysis and developmental status, child protection concerns, and interface with dedicated adolescent health services.

E-health is a complex field, which at the same time should be harmonised and compatible between installations and applications. A strategy to direct planned multi-agent investment is a necessary tool to ensure orderly and efficient progress, focused on clear health benefit objectives. WHO encourages the development of national strategies. In 2012 the WHO and the International Telecommunications Union (ITU) published a national e-Health Strategy Toolkit to offer support to those countries developing an e-health vision and strategy as well as for those where there is a necessity to revitalise available strategies [3]. According to WHO and the ITU, such strategies should be based on national health priorities, available and potential resources, and the current e-health environment. In practice, however, a considerable number of countries are struggling with meeting these requirements. In order to ascertain the currency of national policies, the WHO holds for each country a repository of e-health policies and claims this to be a collection of current national e-health strategies [4].

This paper provides a locally informed view as whether these policies are up-to-date and whether children have been considered in the national e-health strategies. It is based on a study to identify to what extent European countries have defined the health needs of children and adolescents in their national strategies and to what extent ICT solutions have been considered as possible means for the implementation of these strategic goals.

2. Methods

The Horizon 2020 funded project Models Of Child Health Appraised (MOCHA), running from 2015 to 2018, is charged with identifying optimal models of primary care for children, including the role of electronic records to support care delivery [5]. In February 2016, in order assess how well children were supported in national e-health policies, the project carried out an analysis of available national e-health strategies as contained in the WHO depository (http://www.who.int/goe/policies/countries/en/, (access dates 8 and 9/02/16). A key methodological feature of the MOCHA project is the retention in each study country of a part-time Country Agent – a local expert in child health services – who acts as the informant for obtaining data requested by the principal scientists in the project, using local indigenous sources. Questions asked of Country Agents have passed through internal and external scientific scrutiny to confirm their rationale, relevance, and clarity.

To investigate whether children and e-health are considered in national e-health strategies, data were gathered between 21st March and 29th April 2016 through the MOCHA country agent network, thus ensuring local analysis in national languages. The questionnaire was designed as a semi structured survey instrument and asked whether the e-health strategy available in the WHO depository was the current document for each country, and whether it was the only one. MOCHA country agents were asked to list any other relevant e-health document, and to ascertain the presence

of content of each document regarding e-health and children, or the total absence of mention of children.

At the time six out of 30 national e-health strategies were readily available in English with a working link, for 16 countries the documents were available in the countries' own languages only and for eight countries the document was not accessible at all either because a false link was provided or the document was not available at the link provided.

3. Results

Replies were initially received from 27 countries. Eleven out of 27 countries' MOCHA Country Agents replied that the e-health strategy contained in the WHO depository (as of April 2016) was the latest. Twenty said that beside the documents contained in the WHO depository there were other national strategies. Only eleven countries - Cyprus, Germany, Hungary, Ireland, Latvia, Lithuania, Norway, Poland, Portugal, Spain and the UK - mentioned that their countries' e-health strategy contained considerations on children and adolescents. Sixteen replies stated that their national e-health strategy did not consider children and adolescents. Data for 3 countries are still outstanding. Details are given in table 1 and Figure 1.

Table 1: Overview on document currency and context (April 2016) [1,2]

Is the WHO document the latest?		Is the WHO document the only document?	
Yes	No	Yes	No
Bulgaria Cyprus* Finland* Hungary Italy Latvia Lithuania Portugal* Slovakia Spain* Sweden * WHO update since June 2016	Austria* Belgium* Croatia Czech Rep* Estonia Germany Greece* Iceland* Ireland* Luxembourg Malta NL Norway Poland* Romania UK	Finland* Hungary Italy Latvia Romania Slovakia Sweden	Austria* Belgium* Bulgaria Croatia Cyprus* Czech Rep* Estonia Germany Greece* Iceland* Ireland* Lithuania Luxembourg Malta NL Norway Poland* Portugal* Spain* UK
11	16	7	20

[1] Independent of the analysis in this paper, the WHO has updated the repository on national e-health strategies in June 2016. Because field work had been completed in May 2016, it has not yet been possible to update the paper.

Only eleven of 27 countries mention children and adolescents in their national e-health strategy documents, ranging in some cases from mere reflections of data protection concerns, to other countries' comprehensive approaches covering a broad variety of potential areas for the improvement of child primary care through information and communication technology. Whether this weak coverage is due to lack of awarenss of the importance of specific data items and healthcare delivery processes for children, insufficient prioritisation of child healthcare, a lack of appreciation of ICT solutions in the healthcare context, or a lack of resources to sufficiently address challenges linked to a possible implementation, can not be answered at this stage. It is, however, astonishing how little strategic thought is currently spent on the potenial of ICT approaches to make child health care safe and efficient. Clear descriptions of current child health needs, definitions of goals for improvement in child health care, stakeholders to be involved and aligned ICT measures as means of improvement are largely missing.

Figure 1 provides an overview on countries considering children and adolescents in national e-health strategies and those that do not, independent from the actual implementation status of the strategies. This is based on the content analysis of national strategy documents and the MOCHA country agents' evaluation on whether children and adolescents were considered in theory. In cases where national e-health strategies do not refer to the needs and vulnerability of children and adolescents, this has to be considered a potentially serious gap given children's dependence on effective modern systems as being a necessary prerequitie to making child and adolescent health care safe and efficient.

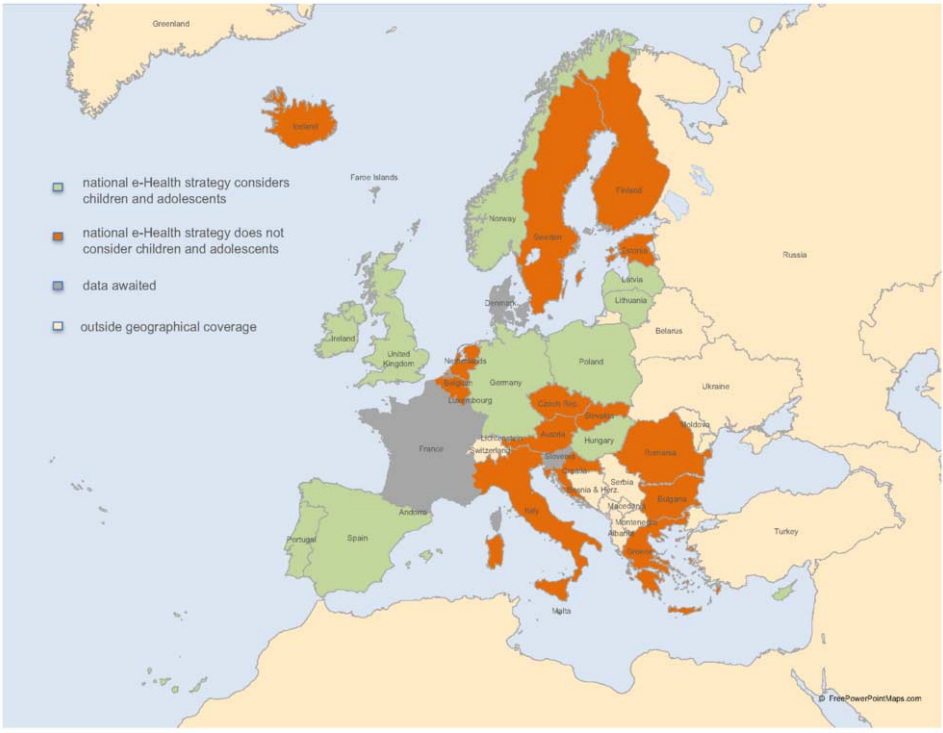


Figure 1. Overview on national e-health strategies

4. Conclusion

Sixteen (59%) of the 27 reporting MOCHA countries fail to give any consideration of children in their national health strategies, and of the 11 (41%) which mention children, seven are in a stage of early adoption of e-health solutions to child health. This underpins how little strategic thought is spent on child health interests and the ICT solutions for more effective and safer child health care, even though IT applications in child health have been a proven cost-effective European innovation for over 50 years [6, 7].

The appropriate development of electronic health records and e-health services for children is severely compromised if under 50% of reporting countries have any mention of the particular needs of children and adolescents, their health, and equitable efficient and effective modern health delivery and monitoring. The findings of this study do not accord with the societal and health system duty of care to children who are dependent on them and cannot advocate for themselves.

This work will be progressed through an inventory of system availability (in hand), and matched to other MOCHA work on patterns of healthcare provision and optimal future models of delivery of children's primary care. Enhanced e-health support will be part of that visioning, and this first policy analysis will help formulate the degree and direction of strategy development needed, as well as initiating discussion on the issues.

Acknowledgements

Acknowledgement is paid to the Country Agents of the MOCHA project as identified on the project website http://www.childhealthservicemodels.eu/partners/). Their contributions ensured that the findings of this study are based on detailed and local indigenous knowledge.

The Models of Child Health Appraised (MOCHA) project is funded by the European Commission through the Horizon 2020 Framework under the grant agreement number: 634201. The sole responsibility for the content of this paper lies with the authors. It does not necessarily reflect the opinion of the European Union, or of the full project. The European Commission is not responsible for any use that may be made of the information contained therein.

References

[1] EU Task Force on eHealth. Redesigning health in Europe for 2020. 2012
[2] M.J. Rigby (2004). Information as the patient's advocate. In M.J. Rigby (Ed.), *Vision and Value in Health Information* (pp. 57-67), Radcliff Medical Press Ltd., Oxon, 2004.
[3] World Health Organisation, International Telecommunication Union. National eHealth Strategy Toolkit. 2012.
[4] World Health Organisation. Directory of eHealth policies 2016 [Available from: http://www.who.int/goe/policies/countries/en/, access date: 08 and 09 February 2016
[5] MOCHA Project www.childhealthservicemodels.eu, access date: 07 March 2016
[6] T. McL. Galloway. Management of Vaccination and Immunization Procedures by Electronic Computer; Medical Officer, 109, 232, 1963.
[7] J. Saunders. Results and Costs of a Computer-assisted Immunization Scheme; British Journal of Social and Preventive Medicine, 24, 187-191, 1970.

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-63

Shared Decision Making via Personal Health Record Technology for Routine Use of Diabetic Youth: A Study Protocol

Selena DAVIS [a,1], Abdul ROUDSARI [a], and Karen L. COURTNEY [a]
[a] *School of Health Information Science, University of Victoria, Victoria, BC, Canada*

Abstract. Engaging patients in the self-management decision-making provides opportunities for positive health outcomes. The process of shared decision making (SDM) is touted as the pinnacle of patient-centred care, yet it has been difficult to implement in practice. Access to tools resulting from the integration of all health data and clinical evidence, and an ease of communications with care providers are needed to engage patients in decision making. Personal health record (PHR) technology is a promising approach for overcoming such barriers. Yet there is a scarcity of studies on system design for SDM via PHR. This paper describes a study protocol to identify functional requirements of PHR for facilitating SDM and factors that would influence the embedding of the proposed system in clinical practice.

Keywords. Personal health records, shared decision making, normalization process theory, mixed methods, self-management, patient engagement

1. Introduction

Today, there is increased interest in health information technology (HIT) interventions that engage patients in decision making as part of self-management. Shared decision making (SDM) has been suggested as an optimal approach to making healthcare decisions and today touted the pinnacle of patient-centred care [1]. SDM is a collaborative process that allows patients and their providers to make decisions together, taking into account the best available evidence and the patient's values and preferences to identify the best strategy at a particular point in time [2]. While there is associated evidence of patient outcomes [3], a few obstacles still slow SDM spread in practice [4].

For patients to be effective and engaged participants in SDM, they require access to their healthcare information and decision support tools. Increasingly, patients are engaged through online, mobile and digital routes [5]. A personal health record (PHR) is an internet-based application that allows patients to access, input, manage and share their health information, access decision support tools and education, and to communicate with care providers [6]. PHRs remain underutilized, but are a major opportunity to improve patient engagement and decision making [7].

To effectively engage patients and support SDM, systems must be designed for that purpose [8]. But using such systems in routine clinical practice is still problematic; a translational gap that Normalization Process Theory (NPT), may be able to address [9].

[1] Corresponding Author: Selena Davis, School of Health Information Science, University of Victoria, PO Box 1700, HSD A202, Victoria, BC V8W 2Y2; Email: daviss@uvic.ca.

Evaluation with NPT focuses on a specific set of activities that bring about the embedding of a complex healthcare practice. NPT describes those determinants that influence the promotion or inhibition of complex interventions and offers a foundation on which the likelihood of successful implementation can be judged [10].

The target population of the study protocol is youth (18-24 years old) with Type 1 diabetes (T1DM). T1DM is the second most common chronic disease in children [11] and an increasing prevalence [12]. While youth can perform diabetes self-management tasks, they still need help making decisions. In terms of engaging diabetic youth in self-management decision making, SDM provides opportunities for positive health outcomes [13]. Still, there are few targeted interventions to support their involvement in SDM [14].

1.1. Objectives

The research objectives are:

- To analyze the functional requirements of an integrated shared decision making–personal health record (*i*SDM-PHR) system
- To explain the factors that promote or inhibit the incorporation of *i*SDM-PHR into routine clinical practice
- To describe the 'normalization potential' of *i*SDM-PHR

1.2. Guiding Theoretical Framework

NPT matches four social processes (the "what") with mechanisms (the "how") and describes the participants' actions which positively affect the embedding of a healthcare intervention. Using *i*SDM-PHR, the sixteen analytical NPT claims are applied (Table 1 adapted from [15][16][17]).

Table 1. NPT Theoretical Framework for SDM via PHR

	Social Process			
Mechanisms	**Meaning and Sense-making work**	**Commitment and Engagement work**	**Enacting work**	**Appraisal work**
Normative restructuring	Understand how SDM via PHR differs from existing practice	Believe it is right for them to be involved	Support and resource SDM via PHR within its social contexts	Identify or measure SDM-PHR benefits and issues
Reworking conventions and group processes	Shared understanding of the purpose and expected benefits	'Buy into' the idea of SDM via PHR and persuade others to participate	Create and communicate knowledge about SDM via PHR	Shared evaluation of contributions to and value of SDM via PHR
Enacting practices	Understand the specific tasks and responsibilities in implementation of SDM via PHR	Willing to drive implementation of SDM via PHR	Operationalize tasks of SDM via PHR and produce outcomes	Evaluate contributions to and value of SDM via PHR
Projecting practices into the future	Understand the value and benefits of SDM via PHR	Commit and contribute to SDM via PHR for sustainability	Allocate roles and responsibilities clearly	Attempts to alter SDM via PHR are made

2. Methods

2.1. iSDM-PHR application

Based on a conceptual framework developed by Davis et al. [18], the design of an *i*SDM-PHR system was developed and explained elsewhere [19]. The proposed *i*SDM-PHR ecosystem is described as an internet-based, electronic health record (EHR) systems environment. It is complemented by autonomous integrated applications whereby data is kept separate from the applications, enabling greater innovation in the applications [20], and built on standards for privacy, security and data exchange, and uses a messaging system for timely, asynchronous interactions . In this way, provider EHR systems and *i*SDM-PHR may interact to exchange relevant data and communicate. Patients access the system anywhere, anytime using any device with internet access. The *i*SDM-PHR functional model maps the elements in the SDM process with a PHR function by patient activity with specific system actions for patients.

2.2. Study Design

The evaluation strategy involves a two-phased, sequential assessment. The protocol is under review by the University of Victoria Human Research Ethics Board. Phase 1 will inform the *i*SDM-PHR functional model via task/function mapping, a user-centred design approach. This process specifies the users' functional requirements of the system. The resultant user-validated, functional model of *i*SDM-PHR will be used in Phase 2.

Phase 2 will assess the 'normalization potential' of the system from multiple user type perspectives. A pre-implementation assessment will be used to identify factors for a successful implementation. Semi-structured interviews will be used to explore participants' opinions of *i*SDM-PHR. Concurrently, the Normalization MeAsure Development (NoMAD), a NPT-based survey instrument [21], will be used to measure implementation processes and predictive relationships between these processes and outcomes. Qualitative and quantitative data will be equally-weighted.

2.2.1. Phase 1 Functional Requirements Evaluation

A purposeful, maximum variance recruitment strategy will be used to gain different perspectives of the requirements of the two target groups: T1DM youth and healthcare providers (physician specialists and certified diabetes educators). The final number of participants will be determined when saturation is reached within data collection; 10-20 participants per target group is anticipated.

Basic descriptive data will be collected including use of technology. Data will be primarily collected via a functional model validation activity and analyses will occur simultaneously. Participants will match a user task with the system function using a task/function matrix, adapted from Maguire [22]. Matching is accomplished when the user identifies a system function for a given SDM task as critical to task, for occasional use, or not useful/applicable to the task. Data collected from each additional participant will be iteratively compared. A functional model will be indicated by 75% of tasks being reliably mapped to functions. As a secondary method of data collection, a brief, semi-structured individual interview will commence in order to clarify and to gain a richer understanding of the participant responses in the functional model validation activity.

2.2.2. Phase 2 Implementation Process Evaluation

Purposeful, maximum variance sampling will be used to assess the 'normalization potential' of *i*SDM-PHR. The three target groups for this phase are T1DM youth, healthcare providers, and organizational providers (responsible for the design, development, implementation or management of EHR systems). Phase 1 participants may participate in Phase 2. Based on similar studies [7][23][24], a study sample of 80 participants (about 20 from each target group) is planned.

Data will be collected concurrently using the following manner: (i) online NoMAD instrument. The NoMAD instrument uses a 5-point Likert scale for 20 items reflecting the full range of normalization processes and 3 items to assess participants' general expectations of the implementation process; (ii) small fixed response survey of practice related outcomes – e.g. "I feel that *i*SDM-PHR would positively impact engagement in self-management decision making?"; and (iii) semi-structured phone interviews that will use an interview guide (Table 2) and be audiotaped and transcribed.

Table 2. Sample Interview Questions

NPT Social Process	Mechanism	Questions/Probes using the NPT theoretical framework
Meaning and Sense-making work	Projecting practices into the future	What do you understand to be the value, benefits and importance of *i*SDM-PHR?
	Normative restructuring	How would you describe *i*SDM-PHR?
Commitment and Engagement work	Normative restructuring	Do you believe it's right to engage in the use of *i*SDM-PHR?
Enactment work	Projecting practices into the future	How does the *i*SDM-PHR affect roles and responsibilities or training needs?
	Reworking conventions and group processes	Will *i*SDM-PHR make people's work easier? Will it impact division of labour, resources, power, and responsibility?
Appraisal work	Enacting practices	How will you judge the value of *i*SDM-PHR in terms of the effects on you?

Demographic, clinical, instrument and survey data will be summarized via descriptive statistics using SPSS to indicate where participants express more positive or negative responses and to inform associations between factors influencing normalization and engagement in decision making. Data from interviews will be analysed with NVivo 11 Pro and coded using a NPT-based coding frame. Salient themes will be grouped to reflect the promoting or inhibiting factors of *i*SDM-PHR into routine practice for diabetic youth. Finally, both quantitative and qualitative data will be merged to a unified whole and analyzed as a composite picture for the purposes of complementarity in the interpretation and description of the outcomes in relation to the research objectives of the study.

3. Discussion

The state of SDM in clinical practice is a question of understanding the operationalization of SDM for patients and providers within today's EHR environment in order for it to become routine. By actively integrating SDM into PHR technology, *i*SDM-PHR may not only improve self-management decision making in diabetic youth, but facilitate improved patient-provider relationships and systematic discussions on further disease management strategies and potential intervention points.

4. Limitation

A non-representative sample does not permit generalization to other populations and cognitive bias cannot be quantified. However, if mixed methods are applied, valid conclusions can be drawn with regard to system design and implementation for use.

References

[1] M. J. Barry, S. Edgman-levitan, and V. Billingham, "Shared Decision Making — The Pinnacle of Patient-Centered Care," *N. Engl. J. Med.*, vol. 366, no. 9, pp. 780–781, 2012.

[2] V. M. Montori, "Shared Decision Making." [Online]. Available: http://shareddecisions.mayoclinic.org/.

[3] L. A. Shay and J. E. Lafata, "Where Is the Evidence? A Systematic Review of Shared Decision Making and Patient Outcomes," *Med. Decis. Mak.*, vol. 35, no. 1, pp. 114–131, Jan. 2015.

[4] F. Légaré and H. Whitteman, "Shared decision-making: examining key elements and barriers to adoption into routine clinical practice," *Health Aff.*, vol. 32, no. 2, pp. 276 – 284, 2013.

[5] C. Corrie and A. Finch, "Expert patients," UK, 2015.

[6] N. Archer and M. Cocosila, "Canadian patient perceptions of electronic personal health records: An empirical investigation," *Commun. Assoc. Inf. Syst.*, vol. 34, pp. 389–406, 2014.

[7] S. Wells, R. Rozenblum, A. Park, M. Dunn, and D. W. Bates, "Personal Health Records for Patients with Chronic Disease," *Appl. Clin. Inform.*, vol. 5, no. 2, pp. 416–429, 2014.

[8] T. Lee, *Information Technology for Patient Empowerment in Healthcare*. Walter de Gruyter GmbH & Co KG., 2015.

[9] E. Murray et al., "Normalisation process theory: a framework for developing, evaluating and implementing complex interventions.," *BMC Med.*, vol. 8, no. 1, p. 63, 2010.

[10] C. May et al., "Evaluating complex interventions and health technologies using normalization process theory: development of a simplified approach and web-enabled toolkit," *BMC Health Serv. Res.*, vol. 11, no. 1, p. 245, 2011.

[11] G. Dovey-Pearce and D. Christie, "Transition in diabetes: young people move on – we should too," *Paediatr. Child Health (Oxford).*, vol. 23, no. 4, pp. 174–179, 2013.

[12] A. M. Sheehan, A. E. While, and I. Coyne, "The experiences and impact of transition from child to adult healthcare services for young people with Type 1 diabetes: a systematic review," *Diabet. Med.*, vol. 32, no. 4, pp. 440–458, Apr. 2015.

[13] D. Koller, N. Khan, and S. Barrett, "Pediatric Perspectives on Diabetes Self-Care: A Process of Achieving Acceptance," *Qual. Health Res.*, vol. 25, no. 2, pp. 264–275, Feb. 2015.

[14] B. Feenstra et al., "Interventions to support children's engagement in health-related decisions: a systematic review.," *BMC Pediatr.*, vol. 14, p. 109, 2014.

[15] C. May, "Agency and implementation: Understanding the embedding of healthcare innovations in practice," *Soc. Sci. Med.*, vol. 78, no. 1, pp. 26–33, Feb. 2013.

[16] K. Gallacher, C. R. May, V. M. Montori, and F. S. Mair, "Understanding Patients' Experiences of Treatment Burden in Chronic Heart Failure Using Normalization Process Theory," *Ann. Fam. Med.*, vol. 9, no. 3, pp. 235–243, May 2011.

[17] F. S. Mair et al., "Factors that promote or inhibit the implementation of e-health systems: an explanatory systematic review," *Bull. World Health Organ.*, vol. 90, no. 5, pp. 357–364, May 2012.

[18] S. Davis et al., "Shared Decision Making Using Personal Health Record Technology: A Scoping Review at The Crossroads", *J. Am. Med. Inform. Assoc.*, 2017. doi:10.1093/jamia/ocw172.

[19] S. Davis, A. Roudsari, and K. L. Courtney, "Designing Personal Health Record Technology for Shared Decision Making", *Studies in Health Technology and Informatics, IOS Press, 234, pp.75-80,* 2017.

[20] P. F. P. Brennan, S. Downs, and G. Casper, "Project HealthDesign: rethinking the power and potential of personal health records," *J. Biomed. Inform.*, vol. 43, no. 5 Suppl, pp. S3–S5, Oct. 2010.

[21] T. L. Finch et al., "NoMAD: Implementation measure based on Normalization Process Theory. [Measurement instrument].," 2015. [Online]. Available: http://www.normalizationprocess.org.

[22] M. Maguire, "Methods to support human-centred design," *Int. J. Hum. Comput. Stud.*, vol. 55, no. 4, pp. 587–634, Oct. 2001.

[23] H. C. Ossebaard, E. R. Seydel, and L. van Gemert-Pijnen, "Online usability and patients with long-term conditions: A mixed-methods approach," *Int. J. Med. Inform.*, vol. 81, no. 6, pp. 374–387, 2012.

[24] D. Gioia and G. Dziadosz, "Adoption of evidence-based practices in community mental health: A mixed-method study of practitioner experience," *Community Ment. Health J.*, vol. 44, no. 5, pp. 347–357, 2008.

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-68

A Medication Reminder Mobile App: Does It Work for Different Age Ranges

Mina FALLAH[a,1] and Mobin YASINI[b]
[a]*School of Allied Medical Science, Tehran University of Medical Sciences,Tehran, Iran*
[b]*DMD santé, Research and development department, Paris, France*

Abstract. Successful medication adherence particularly in elderly with chronic diseases will improve their self-management. Medication reminder systems could be useful to improve this adherence. This study consists of two phases, designing a mobile medical app based on Android platform and then its evaluation. To develop this application, first, the use case scenarios have been hypothesized in partnership with health professionals and patients used to take medications daily. Unified Modeling Language was used to model the use cases. The evaluation was performed with usability testing and efficacy testing. The results show that the app was well accepted both in young people and older adults. Engaging target users and health professionals in the conception and development of a health-related app could have better results in the usability and the efficacy of the app.

Keywords. mHealth, Reminder System, patient adherence,Mobile Applications

1. Introduction

Improving medication adherence particularly in older adults with chronic disorders will enhance their disease management. This would be more serious when the medicines are vital. Most patients occasionally forget to take their medications[1]. Patient adherence to medication is clinically crucial in reducing mortality of serious diseases and total health care costs [2].

Patients with chronic conditions, regardless of age, take generally multiple daily medications with various frequencies. This will increase the risk of medication errors or non-adherence to medication treatment. Non adherence to medication is a complex problem that can lead to exacerbations of chronic health conditions, hospital admissions, and other avoidable health care costs [3].

Smartphones have been rapidly adopted by the general population and now represent a promising technology that can improve health care [4]. The intersection of mobile technology, apps and healthcare is currently in its most dynamic phase. Since the information and communication technology (ICT) develops, a system using mobile phones to support medication-taking will become increasingly necessary as a part of the m-Health (mobile health) system [5].

There are more than 165000 health related apps available in the global markets today [6]. All of these apps are not necessarily “good” ones and studies have shown that many of these applications have some problems concerning content validity,

[1] Mina Fallah, Email: minafallah91@gmail.com; Tel: +989128574373

security, usability, etc. [7]. Unfortunately, the five-star rating scale provided in the app stores is not a reliable assessment method [8]. Some studies in the literature have already demonstrated that older people differ from young people in their perceptions, preferences and usage of mobile technology. There are also differences within the older adult groups regarding mobile technology adoption [9]. Generally, the people using new technologies are young; it seems that the elderly may use mobile applications rarely. However, older adults may have more need to a medication reminder system than the youngsters.

Well-designed applications that are proper for elderly would be adopted by the elderly. We hypothesized that a medication reminder system would be used among all age ranges if it is designed with the participation of its target users and health professionals.

Therefore, in this study, a medication system reminder app has been designed and evaluated by the users from various age ranges.

2. Methods

In the analysis phase, first, we identified the main actor who is the application user. Each user can add one or more patients (medication consumers) and for each patient, the user could allocate various medications. In the first step, the user should identify the patient information including name, gender, and date of birth, height, and weight and phone number (only the name is mandatory).

For visualizing, constructing and, documenting the application, all of the use cases should be recognized. Two IT specialists and two medical doctors and two patients, one above 50 years old and the other under 50 years old, both used to take various medications daily, agreed on the list of use cases and validated the model obtained.

The recognition of the activities of the system was the second step of the development. The activity diagram helps to understand the main business of the system.

Once the application had been developed, one medical doctor and two users tested the application to find bugs and defects.

The evaluation was then performed on the efficiency and the user friendliness of the application. We have created a questionnaire including two questions assessing the function efficacy of the app and 10 usability questions. The usability questions were derived from SUS method [11]. We have adapted the SUS questions for mobile application use. We asked 60 users (30 under 50 years old and 30 over 50 years old) to use the application for at least 10 days and answer to the questionnaire. The evaluators were from both sexes. We have then analyzed the global user satisfaction and the application efficacy and compared the results in the two groups.

3. Result

3.1. Development

The actor list includes the user (the person who uses the application for himself or for others), medication consumer (the person who should take the medication), medication (the medication that should be taken by the patient), patient relative (if user does not take medication, a text massage will be sent to the patient's relative) and trip (Specify a

time limit for providing the medicines in sufficient quantities. Figure 1 shows the use case model.

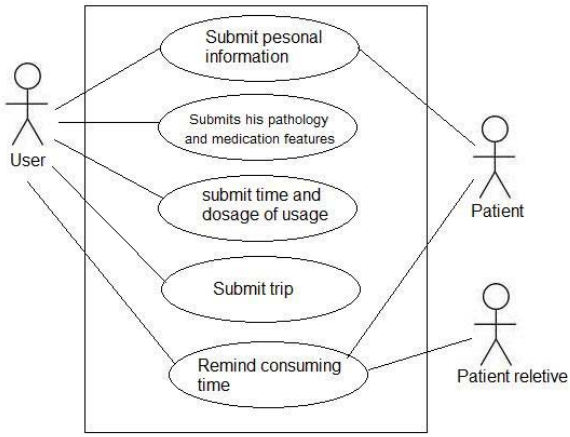


Figure 1. Use case Model

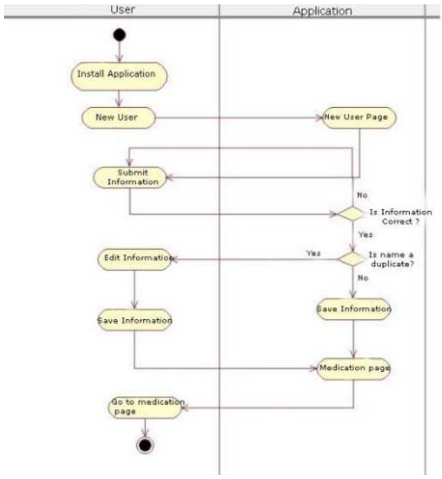


Figure 2. User Creation Activity

The system activities include: user creation, medication creation, consumption on a trip. Figure2 presents an example of activities (the user creation activity).

Our app database contains five tables: User, Alarm, Medication-alarm, User-Medication, and Medication-Info.

3.2. Evaluation

Table 1 shows the efficacy questions and the SUS questions adapted to mobile apps and the mean score for the two groups of users. Both efficacy and usability test results

are promising. We did not find a significant difference in efficacy and global usability between the two groups of study.

Table1. Efficacy and usability evaluation of the application. The data is shown by mean values

Efficacy questions	Under 50	Over 50	*P* value
Compared to the period when I did not use the app, the number of missing medication is reduced (0 for no , not at all and 10 for yes , definitely)	8	8.3	0.2
Compared to the period when I did not use the app, I miss less my medication but I take them sometimes with a delay which is generally more than 1 hour (0 for totally disagree , 10 for totally agree)	2.6	2.9	0.18
Usability questions (0 for totally disagree , 5 for totally agree)			
1. I think that I would like to use this app frequently	4.1	4.6	0.001
2. I found the app unnecessarily complex	1.2	1.4	0.04
3. I thought the app was easy to use	4.5	4.7	0.09
4. I think that I would need the support of a technical person to be able to use this app	1.1	1.7	0.0001
5. I found the various functions in this app were well integrated	4	4.5	0.001
6. I thought there was too much inconsistency in this app	1.3	1.4	0.2
7. I would imagine that most people would learn to use this app very quickly	4.4	4.4	-
8. I found the app very cumbersome to use	1.3	1.4	0.2
9. I felt very confident using the app	4.3	4.5	0.1
10. I needed to learn a lot of things before I could get going with this app	1.2	1.6	0.002
SUS Score	87.8	88.2	0.4

4. Discussion and Conclusion

Medication adherence can affect patient health positively, improve the quality of the relationship between the patients and their health provider, and decrease health resource consumption [12]. Traditional reminder methods remind passively to take medication and are inefficient for complicated regimens.

In this study we've created a medication reminder app that could be appropriate for whom needing support with his drug regimens. The app can engage the medication consumer's entourage and the consumer should not carry a separate reminder device. The comparison between the two groups of study shows that the application was well accepted in the elderly and the young generation both in efficacy and usability scores.

A number of researches studied the use of smartphones in the medical settings. One systematic review study found that the short-term effectiveness of electronic reminders, especially Short Message Service (SMS) text messaging reminders and internet interventions could enhance medication adherence of the patients [12, 13]. The

results of our study corroborate the findings of other studies and show that the age of users does not affect the usability and efficacy of the app. We have found significant differences in some usability aspects and contrary to our expectations; they were better noted by the older group. This may be due to the comparison that the young people do when assessing an app. The young people have seen and tested more apps than the older group. Therefore, the expectations of this group may be higher.

One of the limitations of this study may be our small sample size. Another limitation would be the fact that we have evaluated our own app and the results may not be generalizable to other solutions. However, we believe that if the target users and health professionals participate in the conception and the development of the app, it will obtain a good level of acceptance[8]. Other usability and efficacy evaluation methods could be used in the future to validate these results. We considered "age" as a primary factor in this study. However, other factors including the knowledge, illness status and perceived support could be integrated in the future revisions of this study. The future app may have a verity of features including medication information, multi-platform functionality, and interoperability with the electronic patient records and particularly order entry systems.

References

[1] Eagleton JM, Walker FS, Barber ND. An investigation into patient compliance with hospital discharge medication in a local population. International Journal of Pharmacy Practice. 1993;2(2):107-10.
[2] Simpson SH, Eurich DT, Majumdar SR, Padwal RS, Tsuyuki RT, Varney J, et al. A meta-analysis of the association between adherence to drug therapy and mortality. Bmj. 2006;333(7557):15.
[3] Billingsley L, Carruth A. Use of Technology to Promote Effective Medication Adherence. Journal of continuing education in nursing. 2015;46(8):340-2.
[4] Anglada-Martinez H, Rovira-Illamola M, Martin-Conde M, Sotoca-Momblona JM, Codina-Jane C. mHealth intervention to improve medication management in chronically ill patients: analysis of the recruitment process. Postgraduate medicine. 2016;128(4):427-31.
[5] Boulos MN, Wheeler S, Tavares C, Jones R. How smartphones are changing the face of mobile and participatory healthcare: an overview, with example from eCAALYX. Biomedical engineering online. 2011;10(1):24.
[6] Powell AC, Torous J, Chan S, Raynor GS, Shwarts E, Shanahan M, et al. Interrater Reliability of mHealth App Rating Measures: Analysis of Top Depression and Smoking Cessation Apps. JMIR mHealth uHealth. 2016;4(1):e15.
[7] Boulos MN, Brewer AC, Karimkhani C, Buller DB, Dellavalle RP. Mobile medical and health apps: state of the art, concerns, regulatory control and certification. Online journal of public health informatics. 2014;5(3):229.
[8] Yasini M, Marchand G. Mobile Health Applications, in the Absence of an Authentic Regulation, Does the Usability Score Correlate with a Better Medical Reliability? Studies in health technology and informatics. 2015;216:127-31.
[9] Vicente P, Lopes I. Attitudes of older mobile phone users towards mobile phones. Communications. 2016;41(1):71-86.
[10] Linn AJ, Vervloet M, van Dijk L, Smit EG, Van Weert JC. Effects of eHealth interventions on medication adherence: a systematic review of the literature. Journal of medical Internet research. 2011;13(4):e103.
[11] Harbig P, Barat I, Damsgaard EM. Suitability of an electronic reminder device for measuring drug adherence in elderly patients with complex medication. Journal of telemedicine and telecare. 2012;18(6):352-6.
[12] Dayer L, Heldenbrand S, Anderson P, Gubbins PO, Martin BC. Smartphone medication adherence apps: Potential benefits to patients and providers. Journal of the American Pharmacists Association : JAPhA. 2013;53(2):172-81.
[13] Su W-C, Chih M-Y. Is More eHealth System Use Better for Cancer Patients and Family Caregivers? A Literature Review. American Medical Informatics Association Poster; November 12-16, 2016; IL.

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-73

Internet of Things in Health Trends Through Bibliometrics and Text Mining

Stathis Th. KONSTANTINIDIS[a,1], Antonis BILLIS[b], Heather WHARRAD[a], Panagiotis D. BAMIDIS[b]
[a] *School of Health Sciences, The University of Nottingham, United Kingdom*
[b] *Medical School, Aristotle University of Thessaloniki, Greece*

Abstract. Recently a new buzzword has slowly but surely emerged, namely the Internet of Things (IoT). The importance of IoT is identified worldwide both by organisations and governments and the scientific community with an incremental number of publications during the last few years. IoT in Health is one of the main pillars of this evolution, but limited research has been performed on future visions and trends. Thus, in this study we investigate the longitudinal trends of Internet of Things in Health through bibliometrics and use of text mining. Seven hundred seventy eight (778) articles were retrieved form The Web of Science database from 1998 to 2016. The publications are grouped into thirty (30) clusters based on abstract text analysis resulting into some eight (8) trends of IoT in Health. Research in this field is obviously obtaining a worldwide character with specific trends, which are worth delineating to be in favour of some areas.

Keywords. scientometrics, ubiquitous health, pervasive health, text data mining

1. Introduction

A new buzzword has come into the foreseen recently, namely Internet of Things. Kevin Ashton claim the first use of term Internet of Things (IoT) in 1999 [1] linking the idea of RFID in a supply chain. A lot of definitions are attempted with the Oxford dictionary defining IoT as "*the interconnection via the Internet of computing devices embedded in everyday objects, enabling them to send and receive data"*. The IEEE Internet Initiative definition as explained in [2] distinguishes between low and high complexity, thus it established separate definitions. Following the work of pervasive or ubiquitous computing, the Internet of Things in Healthcare is one of the key focus areas with examples including use in Active and Healthy Aging environments [3], "personalised preventative health coaches" [4], full body exergames to mobile devices [5] and others.

Bibliometric analysis provides a summary for research reported in scientific literature enabling researchers to generate quantitative information from existing data [6]. Text data mining can complement the bibliometric analysis and fulfil the need for faster content analysis and categorisation [7]. Text data mining or text mining involves information retrieval, text analysis, information extraction, clustering, categorization, visualization, database technology, machine learning and data mining [8].

[1] Corresponding author, DICE Research Group & HELM Team, School of Health Sciences, The University of Nottingham, Nottingham, NG7 2UH, UK; E-mail: Stathis.Konstantinidis@nottingham.ac.uk.

A recent study for Internet of Things using bibliometrics revealed [9] an increased number of publications in the last 5 years, while emerging and established research clusters were identified. Furthermore, there are a number of organisations which foresaw the rapid expansion of Internet of Things. The US National Intelligence Council include it in 2008 in a list of six "Disruptive Civil Technologies" with potential impact on US national power [10]. The UK government sees IoT as a transformative development which has enormous potential to change all of our lives [11] and foresees the potential areas for development of IoT in Healthcare to be prevention and early identification, research and tailored healthcare. The EU state that IoT represents the next step towards the digitisation of our society and economy and by 2020 the market size of IoT expected to exceed one trillion Euros [12].

Different organisations have different research priorities around Internet of Things in Health. Reviews on Internet of Things aiming to set the vision and identify the application and concepts have been made [13], [14], as well as an analysis via rigorous bibliometric and network analytics [9]. However, to the best of our knowledge there is no current research identifying the Trends of Internet of Things in Health through bibliometrics and text mining. This paper taking into consideration the field diversity in healthcare aims to answer the following questions: What are the IoT in Health bibliometrics? What are the themes in IoT in Health research? Are there any trends for IoT in Health research? What can be envisaged in the future in this domain?

2. Methods

Inspired by the work of Hung [7] we followed a similar methodology.

Data Collection: We chose the Web of Science as the source database since includes journals with highest impact in science and it is a bibliometric database which enables detailed bibliometric analysis. The following query was used to identify the relevant papers: *TS=("Internet of Things" OR IoT) AND (ehealth OR health* OR medic* OR nurs*) NOT (impairment oriented-training OR Integrated outpatient treatment OR Immunotech OR intravenous injection of endotoxin)* in order to include the term "Internet of Things" or the acronym IoT and one or more of words which relates IoT with health (e.g. ehealth, health, healthcare, medicine, medical, nurse, nursing, etc.) in the Abstract, Title and/or Keywords fields of a record. The acronym "iot" is used for different health related terms which excluded from the results. A total of 778 papers were retrieved. The search period was set from 1st of Jan 1998 to 30th Sep 2016. Despite the fact that not all the 2016 papers included, the latest papers might change the scene on the trends of internet of things in Health.

Data Analysis: The bibliometric data generated by the Web of Science downloaded locally and a set of bibliometric indicators were extracted, which describe the IoT in Health. BibExcel used to calculate co-occurrences of countries per paper and Pajek to visualise it. Text mining and clustering analysis performed using WordStat. The algorithm used is based on hierarchical clustering of key words (a text mining approach for automatic taxonomy generation and text categorisation), using as similarity metric the Adjusted Phi coefficient, a measure of association for two binary variables. In addition, the clustering method that was chosen was based on co-occurrence profiles (Second Order Clustering), considering that two keywords are close to each other, not necessarily because they co-occur but because they both occur in similar environments. Pre-processing involved stemming and exclusion of common words and phrases.

3. Results

Publication Data Collection: Figure 1(a) summarises the number of publication between 1998 and 2016. Publications of 2016 are included as they appear in Web of Science until September. The solid line represents the number of articles, while the dashed represents the moving average trend line. As the figure reveals the publications of IoT in Health are continuously growing. Document Type: The majority of the published literature in IoT in Health are proceedings papers while less than half are articles in journals. Twenty (20) are reviews while editorials, book chapters and abstracts are also present (Figure 1(b)).

Figure 1. IoT in Health : a) publication trends ; b) document types

The most prolific countries based on authors' affiliations are People's Republic of China (134) followed by USA (113), while UK (63), India (63), Spain (43) and South Korea (42) follow (Figure 2(a)). However, research is being conducted worldwide as figure 2(a) reveals with exceptions of Africa, a part of South America and middle Asia. The highest collaborations between countries is between People's Republic of China and USA, and as figure 2(b) depicts high cross-country collaboration exists. The most prolific source titles are LNCS (21), Sensors (20), Applied mechanics and materials (16) and Procedia Computer Science (15), while the 778 papers are published in 679 different sources.

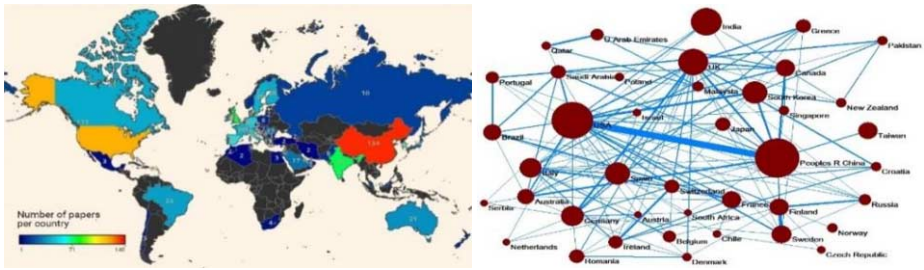


Figure 2. a) World research on IoT in Health based on authors' affiliation. b) Collaboration between countries based on authors' affiliations using BibExcel and Pajek.

Text Mining analysis through automatic clustering analysis created a dendrogram of 30 key term clusters. The 30 clusters interpret by two domain experts. Clusters with less than 5 articles and one cluster with noise (papers that were irrelevant to the topic) were dropped resulting to 745 articles. Since most of the terms exist in more than one article, frequency of cluster terms per article was used, to assign each article into a single cluster. As a result, a total number of eight (8) categories were formed by

combining several clusters together. As a further step, publishing trends were calculated for each category. Time trends are depicted in the Figure 3(b).

	Internet of Things in Health Trends	Articles
C1	Systems /Services design and implementation	258
C2	Communication/Connectivity Protocols & Algorithms	151
C3	Industrial potential of IoT	13
C4	Data Science analysis, storage and connectivity	92
C5	Quality Management and Privacy	23
C6	Efficiency and Cost of Application	172
C7	Smart Cities	15
C8	Ambient Assistive Living & Active Healthy Aging	21

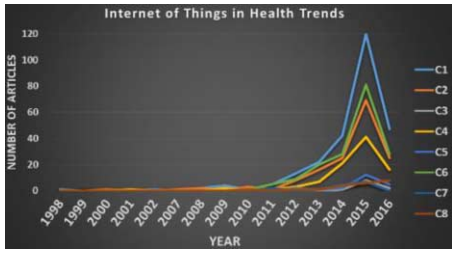


Figure 3. a) Trends identified and number of articles per trend for the selected period (Jan 1998 – Sep 2016). b) Time trends of articles identifying the growing interest of trends (number of articles per year).

The distribution of articles in C1, C2 and C6 trends follow the general worldwide research pattern, while C3 – "Industrial potential of IoT" trend apart from the 2 dominant Countries (Peoples Republic of China and USA (40% of the articles)) articles comes mainly from European countries (46% of the articles). C8 – "Ambient Assistive Living & Active Healthy Aging" is led by Greece, with the rest of the research spread worldwide. C4 – "Data Sciences analysis, storage and connectivity" which includes research on Big data, cloud computing and semantic web, is one of the few trends where Peoples Republic of China are not leading. Instead USA (20% of the articles), UK (15.7% of the articles) and India (10% of the articles) are the origin countries of this research topic. The C5- "Quality management and privacy" topic authoring country origin is led by European countries (43 % of articles), followed by Asia (30%) and North America (20%). Similar pattern applies for C7-"Smart Cities" with Europe to author 52.4% of the total articles followed by North America (23.8%) and Asia (20%).

4. Discussion and Conclusion

As the term 'Internet of Things' was claimed to be coined in 1999 we have selected the time range of the queries to be between 1998 and 2016. The growth of published research started in 2010, while a higher increase occurred after 2014 as Figure 1(a) and 3(a) depict. Since articles for 2016 are not included in full due to the time the search took place, the decreasing curve in the number of articles after 2015 is expected. The authors selected to include articles authored in 2016, since those formed differently the trends and allowed the latest focus of the published articles to be included.

The increased number of proceedings papers in comparison with journal articles (Figure 1(b)) might be an indicator for the rapid expansion of the field and the need for presenting early results and concepts, despite the fact that proceedings age faster they play a particularly important role in computer sciences (about 20% of the references)[15].

The institutional country of the authors reveals the countries where relevant IoT in Health research occurs (Figure 2 (a)) with the two dominant countries (Peoples Republic of China and USA) to have also the highest rate on collaboration between them through co-authorships (Figure 2(b)). As can been seen from Figure 2 cross-country collaborations are essential for this emerging field.

Some of the trends include more articles than others (Figure 3(a)). This could be linked with previous knowledge (e.g. C1, C2, C6) but also with research fields expansions such as big data and semantic web (C4), and the Active and Healthy Aging (C8). The time trends of articles identify a growing interest on the recognised topics. According to Figure 3(b), topics such as IoT industry potential (C3), quality management and privacy (C5) and smart cities (C7) are recently emerging trends since their numbers are increasing since 2014, while other more mature trends include ambient assisted living area (C8), systems design & implementation (C1), communication protocols (C2), data sciences (C4) and cost of implementation (C6), dating back in 2013 and 2011.

In this paper we formed eight (8) trends in Internet of Things in Health research through a bibliometric analysis. Research is worldwide with specific trends seen in certain countries. The findings of the study are limited and do not intended to be exclusive. This study might have missed articles in other scholar databases or articles that are in the process of publication. Different sets of keywords used in the search might generate different results that influence the trends of IoT in Health. This classification is by no means exhaustive and different classifications could be used. It is nevertheless already didactic from this piece of research that there exists already preferred domains of IoT applications which look not only promising for researchers but for facing societal challenges like active and healthy ageing as well.

References

[1] K. Ashton, “That ‘Internet of Things’ Thing,” FRID J., 2009.
[2] R. Minerva, A. Biru, and D. Rotondi, “Towards a definition of the Internet of Things (IoT),” IEEE Internet Initiative. p.86, 2015.
[3] E. I. Konstantinidis, G. Bamparopoulos, A. Billis, and P. D. Bamidis, “Internet of things for an age-friendly healthcare.,” Stud. Health Technol. Inform., vol. 210, pp. 587–91, 2015.
[4] D. V Dimitrov, “Medical Internet of Things and Big Data in Healthcare.,” Healthc. Inform. Res., vol. 22, no. 3, pp. 156–63, Jul. 2016.
[5] E. I. Konstantinidis, G. Bamparopoulos, and P. D. Bamidis, “Transfering full body exergames from desktop applications to mobile devices: The role of the Internet of Things,” in IMCL2015, 2015, pp. 254–258.
[6] M. Thelwall, “Bibliometrics to webometrics,” J. Inf. Sci., vol. 34, no. 4, pp. 605–621, Jun. 2008.
[7] J. Hung, “Trends of e-learning research from 2000 to 2008: Use of text mining and bibliometrics,” Br. J. Educ. Technol., vol. 43, no. 1, pp. 5–16, Jan. 2012.
[8] A.-H. Tan, “Text Mining: The state of the art and the challenges,” in Proceedings of the PAKDD 1999 Workshop on Knowledge Disocovery from Advanced Databases, 1999, vol. 8, pp. 65–70.
[9] D. Mishra, A. Gunasekaran, S. J. Childe, et al, “Vision, applications and future challenges of Internet of Things,” Ind. Manag. Data Syst., vol. 116, no. 7, pp. 1331–1355, 2016.
[10] The National Intelligence Council and SRI Consulting Business Intelligence, “Disruptive Civil Technologies Six Technologies With Potential Impacts on US Interests Out to 2025,” 2008.
[11] The Government Office for Science, “The IOT: making the most of the Second Digital Revolution,” 2014.
[12] S. Aguzzi, D. Bradshaw, et al, “Definition of a Research and Innovation Policy Leveraging Cloud Computing and IoT Combination; A study prepared for the European Commission DG Communications Networks , Content & Technology - Contract number: 30-CE-0608166/00-79,” 2014.
[13] S. Madakam, R. Ramaswamy, and S. Tripathi, “Internet of Things (IoT): A Literature Review,” J. Comput. Commun., vol. 3, no. 3, pp. 164–173, 2015.
[14] G. Russo, B. Marsigalia, et al, “Exploring regulations and scope of the Internet of Things in contemporary companies: a first literature analysis” J. Innov. Entrep., vol. 4, no. 1, pp. 1–13, 2015.
[15] L. Cynthia, L. Vincent, and É. Archambault, “Conference proceedings as a source of scientific information: A bibliometric analysis” J. Am. Soc. Inf. Sci. Technol., vol. 59, no. 11, pp. 1776–1784, 2008.

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-78

Developing the Safety Case for MediPi: An Open-Source Platform for Self Management

Andrew CARR [a], Damian MURPHY [a], Ian DUGDALE [a], Anne DYSON [a], Ibrahim HABLI [b,i] and Richard ROBINSON
[a] *NHS Digital, UK*
[b] *University of York, UK*

Abstract. mHealth and Telehealth technologies are increasingly used to provide personalised, interactive and timely access to health data, thereby helping patients take a more active role in their care process. However, similar to any intervention, the use of these technologies has to be assured to justify that they do not compromise patient safety. In this paper, we discuss the development of a safety case for MediPi; a research prototype for a low-cost open-source digital platform that collects physiological data from patients, at home, and makes it available to decision-support systems used by clinicians. We identify potential hazardous failures associated with the use of MediPi and examine current risk controls. We also explore the modular structure of the overall safety case of the platform. We conclude with a discussion of patient safety challenges related to the unsupervised nature of the care setting and the use of commercial off-the-shelf personal devices.

Keywords. mHealth, telehealth, open source, patient safety, safety case.

1. Introduction

Mobile Health (mHealth) and Telehealth technologies are considered two central digital solutions for enabling patient-centred care [1]. By offering timely, personalised and interactive access to health data and services, a primary aim is to empower patients to take a more engaged role in their care process and improve the quality of the coordination of care between patients and carers. Importantly, the active participation of the patients makes them providers of context- and person-specific health data rather than mere consumers of generic health services.

In particular, the number of mHealth apps that support self-management and self-monitoring has increased significantly. It is estimated that, by 2017, 65% of these apps will focus on capturing and communicating data to measure patient conditions, with emphasis on managing chronic diseases [1]. Within the UK, the telehealth market is expected to grow by 13% (Compound Annual Growth Rate) [2] per annum in the next few years with estimated yearly average costs at £2,000 per patient [3]. This sharp increase is attributed to advances in smart mobile phones, combined with reduced cost of ownership and ease of access to apps and wearable devices.

However, claims concerning the benefits of these digital technologies have to be supported by the necessary evidence of efficacy, cost-effectiveness, usability, safety and

[i] Corresponding author: Department of Computer Science, University of York, York, YO10 5GH, UK; E-mail: ibrahim.habli@york.ac.uk

security. In particular, the potential of these systems to improve or compromise patient safety remains an open question [4]. This adds to the overarching challenge concerning assuring patient safety in the community, as highlighted by Vincent and Amalberti, "*patient safety in the home has not been addressed in a systematic manner*" [5]. For example, remote digital communication between the clinician and patient can increase the risk of missing subtle conditions that are likely to be detected in person [6]. Further, a wider technological issue relates to the reliability of publicly available mHealth apps. A recent assessment of 46 mHealth apps for calculating insulin dose concluded that the majority of these apps lack protection against incorrect or inappropriate dose recommendations that put users at risk of catastrophic overdose [7].

Needless to say, our objective is not to undermine the role of mHealth and Telehealth technologies. Rather, it is to highlight potential patient safety risks posed by these technologies and the need to justify and establish control measures that detect and mitigate these risks. That is, similar to any intervention, mHealth and Telehealth technologies are expected to undergo evaluation and produce assurance evidence that is *proportionate* to the criticality of these technologies to patient safety. Unfortunately, current regulations and guidelines are still evolving and the characteristics of such assurance evidence for mHealth solutions are hard to determine.

In the UK, the Personalised Health and Care 2020 Plan makes a commitment to "*working in the open and ensuring all new source code is open and reusable, and published under appropriate licences, unless there is a convincing reason not to*" [8]. In this paper, we examine how this notion of openness can be extended to safety assurance by augmenting open-source software artefacts with open safety cases. Such open and publically available safety cases comprise specific assurance claims about the open-source technology and the argument and evidence to support these claims.

As a use case to explore the assurance of mHealth, we focus on MediPi, a prototype for a low-cost open-source digital platform that collects physiological data from patients, at home, via physical devices (e.g. oximeters and blood-pressure cuffs) and makes it available to clinicians. The system supports the return of information from clinicians to the patient. The development of MediPi is clinically led and follows an agile lifecycle. It aims to provide a generic and open-source platform for mHealth solutions that can be customised for self-management for different health conditions.

2. Methods

2.1. Setting

In the current prototype, MediPi collects and transmits data as-is with no modification or filtering. Any clinical interpretation of, and reaction to, the information are exclusively the responsibility of the clinicians. The MediPi platform does not provide any direct treatment or diagnosis. A high-level model of the MediPi architecture is shown in **Figure 1**. *Devices* are commercial off-the-shelf USB- or Bluetooth-enabled physiological measurement hardware. Data is collected by the *MediPi Patient Interface* software and transmitted to a *Concentrator,* which makes it available to *Clinical Systems*. The Clinical Systems and Devices are out of scope for the MediPi platform. The Concentrator holds all data collected from the patients, in a raw form, and provides a query mechanism for Clinical Systems, which are responsible for any processing and raising notifications. The Concentrator will forward these notifications to the patient.

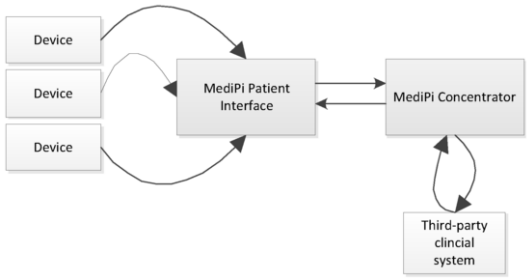


Figure 1. MediPi Solution Architecture

MediPi has a multi-layered approach to security. All connections between MediPi entities occur behind a VPN hosted by NHS Digital. All messaging 'in transit' is mutually authenticated (2-way SSL) using MediPi Device certificates and 'at rest' is encrypted and signed using MediPi Patient certificates. Asymmetric encryption allows access to only intended recipients and the signature ensures the identity of the sending party. Although for the purposes of the current pilot study Raspberry Pi is used as the main hardware platform, there is no requirement for this to be the case in further deployments. MediPi is an open-source Java solution and is platform agnostic. The source code and design are publically available via GitHub[ii], which allows third parties to access the code for their own implementation. A port to Android is under way.

2.2. Approach

We performed a preliminary safety evaluation of MediPi by following the requirements for risk assessment defined in the NHS Digital safety standard for Health IT development [9]. This standard follows the safety principles established for medical devices. The risk assessment of MediPi generated two primary artefacts: Hazard Log and Clinical Safety Case Report. The Hazard Log is a mechanism for recording the on-going identification, analysis and resolution of system hazards. The Clinical Safety Case Report documents an argument, supported by evidence, for why the system is safe for a given application in a given environment.

For hazard analysis, we conducted a *What-If Analysis*, focusing on four categories of failure: *Omission*, *Commission*, *Timing* and *Incorrect,* following the guidewords defined in the Software Hazard Analysis and Resolution in Design (SHARD) technique [10]. For the safety case report, the safety argument was depicted graphically using the Goal Structuring Notation (GSN) [11]. GSN captures the individual elements of the safety argument, e.g. claims, context and evidence, and the relationships that exist between these elements. Both the Hazard Log and the Clinical Safety Case Report will be made publically available alongside the MediPi source code and design specification.

3. Results

Nine potential hazardous failures were identified and analysed **(Table 1)**. An additional generic failure was also identified that concerns patient misidentification. The primary output of MediPi is information, in the form of clinical notifications (e.g. from a GP) or

[ii] http://rprobinson.github.io/MediPi and www.medipi.org

prompts (e.g. periodic reminders to measure blood pressure). As such, MediPi cannot *directly* cause physical harm. That is, hazards in this context are actions or inactions *by the users* to which MediPi notifications and prompts could contribute. For example, the user might not take a blood pressure measurement at the correct time (i.e. *inaction* or *omission* event) due to the lack of a prompt by MediPi.

Table 1. Identified Hazardous Failures

	Clinical Notifications	Clinical Prompts
Omission	MediPi does not present clinical notification	MediPi fails to provide a prompt
Commission	MediPi presents clinical notification when not required	MediPi provides a prompt when not required
Timing	MediPi presents clinical notification later than required	MediPi provides prompts earlier than required MediPi provides a prompt later than required
Value	MediPi presents incorrect clinical notification	MediPi provides an incorrect prompt

For each hazardous failure shown in **Table 1**, existing risk controls were examined and additional mitigations were identified in order to reduce the likelihood/severity associated with these failures. These mitigations specify safety requirements that have to be satisfied by the next design phase (e.g. further redundancy) or through the deployment of specific measures in the care setting (e.g. user training).

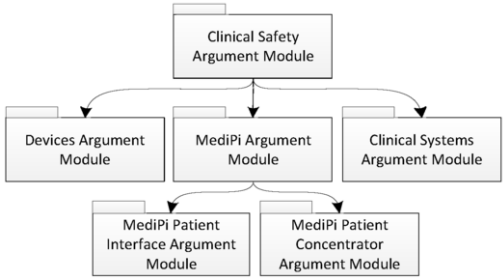


Figure 2. Overall Modular Safety Case

Concerning the safety case, modularity in the MediPi platform, e.g. separate modules for devices, Concentrator and Clinical Systems, lends itself to modularity in the safety reasoning, i.e. self-contained safety arguments for the individual system modules. As such, the overall safety case for MediPi is modular, organised based on interdependent argument modules. These are summarised as follows **(Figure 2)**:

- *Clinical Safety Argument Module*: defines a hazard-directed argument concerning how the risk of identified care hazards has been mitigated;
- *Devices Argument Module*: captures assurance about the behaviour of the commercial off-the-shelf devices, e.g. reliability of the measurements;
- *MediPi Argument Module*: provides a justification concerning two aspects of the platform, Patient Interface and Concentrator, focusing on how the hazardous failures identified in **Table 1** have been mitigated;
- *Clinical Systems Argument Module*: captures assurance concerning the contributions to care hazards made by the clinical inputs.

In particular, the *Clinical Safety Argument Module* is context-specific and as such it is expected to vary to justify the safety of the specific characteristics of the different care settings and system configurations. The correspondence between the compositional design of the MediPi solution and the modular organisation of the safety case is intended to reduce the effort involved in reusing both the solution and its associated safety case without the need to perform the safety assurance from scratch for each deployment.

4. Discussion and Conclusions

The preliminary safety case for the MediPi platform highlighted a set of assumptions that are the subject for further investigation. Firstly, concerning the care setting, each deployment of the platform is customised to meet the specific clinical needs, social setting and personal preferences of the patients. Applicable hazards and their associated levels of risks are expected to vary across the different settings. As such, the safety case for MediPi is considered dynamic and will evolve to provide assurance that is specific to each care setting and cohort of patients. For example, in the current pilot system, physiological measurement devices are read-only and raise no alerts. No apparatus is implemented in the system for treating the patient. Clinical staff members make all decisions. However it is not hard to envisage an evolved version that did perform some control over measurement devices, performed active monitoring, and was able to make decisions or assist the patient in making decisions for themselves. This would lead to an expectation of greater safety rigour in the development of the MediPi platform (*MediPi Argument Module*) and associated clinical systems and the choice of devices.

Secondly, safety cases for systems used in hospital settings tend to appeal to the availability, professionalism and competency of a clinician, i.e. the clinician is in the loop. For an mHealth solution such as MediPi, little credit can be given to the user as a risk control, e.g. the ability of a patient to detect invalid but plausible data is limited. As such, in the current pilot deployment, MediPi is not used to communicate any critical clinical notifications. Thirdly, tradeoffs between technological risks, clinical risks and economic constraints have to be explicitly considered and justified, particularly for cases where the clinical benefits might outweigh residual technological risks. That is, high levels of rigour will not be required for cases where patient safety risk is low or further risk reduction is grossly disproportionate to the benefits gained.

Finally, as with any safety case, no conclusive evidence can be provided concerning the safety of the system. Publically-available and open safety cases, particularly for open-source systems, should benefit from the review of the community, thereby helping to improve the assurance of these systems and promote a shared and open learning culture around patient safety and digital health.

References

[1] WHO Regional Office for Europe, *From innovation to implementation – eHealth in the WHO European Region*, World Health Organization, Copenhagen, 2016.
[2] Office for Life Sciences, Digital Health in the UK: An Industry Study, 2015.
[3] C. Henderson, et al, Cost Effectiveness of Telehealth for Patients with Long Term conditions: Nested Economic Evaluation in a Pragmatic, Cluster Randomised Controlled Trial. *BMJ* (2013), 346:f1035.
[4] L.T. Lorchan & J.C. Wyatt, mHealth and Mobile Medical Apps: A Framework to Assess Risk and Promote Safer Use. *Journal of medical Internet research* 16.9 (2014): e210.
[5] C. Vincent & R. Amalberti, *Safer Healthcare. Strategies for the Real World*, The Health Foundation/Springer Open, London, 2016.
[6] S. Agboola, D. Bates, & J. Kvedar. Digital Health and Patient Safety, *JAMA* 315.16 (2016): 1697-1698.
[7] K. Huckvale, et al. Smartphone Apps for Calculating Insulin Dose: A Systematic Assessment. *BMC medicine* 13.1 (2015): 1.
[8] Department of Health and National Information Board, *Personalised Health and Care 2020*, NIB, 2014.
[9] NHS Digital. SCCI0129, *Clinical Risk Management: its Application in the Manufacture of HIT Systems – Implementation Guidance*. Standardisation Committee for Care Information. 2016.
[10] D.J. Pumfrey, *The Principled Design of Computer System Safety Analyses*. Diss. University of York, 99.
[11] GSN Community Standard, Version 1, http://www.goalstructuringnotation.info [Accessed 26/10/2016].

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-83

UK Health and Social Care Case Studies: Iterative Technology Development

Adie BLANCHARD[a], Laura GILBERT[a] and Tom DAWSON[a1]
[a] *Rescon Ltd., United Kingdom*

Abstract. As a result of increasing demand in the face of reducing resources, technology has been implemented in many social and health care services to improve service efficiency. This paper outlines the experiences of deploying a 'Software as a Service' application in the UK social and health care sectors. The case studies demonstrate that every implementation is different, and unique to each organisation. Technology design and integration can be facilitated by ongoing engagement and collaboration with all stakeholders, flexible design, and attention to interoperability to suit services and their workflows.

Keywords. Technology, integration, health care, social care, software as a service

1. Introduction

Health and social care systems worldwide are facing increasing pressures, driven by an ageing population and increasing prevalence of long term conditions [1]. As a result, there has been a pressing need to enhance service efficiency and deliver high quality care at a lower cost [2]. As this strain increases, technology will play a greater role in the management of health and wellbeing [3].

Technology has been implemented in many services to help reduce the gap between supply and increasing demand [4], providing an opportunity to improve services and outcomes at a reduced cost [5]. Technology can provide many advantages in health and social care including improved efficiency, quality of care, health outcomes, and provision of new services [6].

With substantial advances in health technology, and increase in funding set aside for future digital projects [7], there is a growing need to better understand the development requirements of technology to integrate with existing services to improve uptake and longevity [8]. Research has identified the importance of ongoing engagement with all stakeholders [4], [9], and for organisations to be aware of the resource cost of adopting a new technology [10]. It is advantageous to have adaptable technologies to meet changing needs and requirements of an organisation [9], and interoperability is vital [11].

Whilst there is extensive research into factors surrounding technology design and implementation, there are few practical case studies that provide real world implementation guidance. This paper will discuss the implementation of a Software as a Service (SaaS) application to manage health and wellbeing in a number of different case studies.

[1] Corresponding Author, The Kilns, Penn Croft Farm, Crondall, Hampshire, UK. Email: tom.dawson@rescontechnologies.com

2. Case Study – Software as a Service

Lincus is a SaaS tool for recording and monitoring health and wellbeing information, providing both self-care and shared care capabilities. Lincus allows individuals to quantify and record subjective and objective measures of health and wellbeing using picture based surveys.

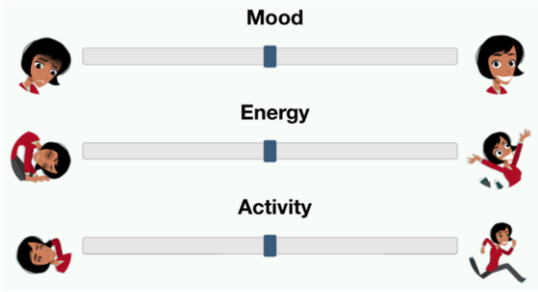


Figure 1. Lincus survey.

Events and/or interventions can be logged, enabling the impact they may have had on health and wellbeing to be identified. Nutrition, physical activity, and clinical measurements can also be recorded. Over time, this provides a history of health and wellbeing, which is communicated to stakeholders using a variety of data visualisations.

Lincus pilot trials have demonstrated usability across a range of user groups. System users have reported numerous benefits as secondary outcomes, including improved health and wellbeing, identification of previously undiagnosed conditions, and enhanced engagement and communication with service providers.

2.1. Case Study 1

Lincus was first piloted in 2013, after considerable co-development with stakeholders, as a usability trial with a service that supports those with multiple and complex needs at risk of homelessness [12]. Initial co-development identified that the system would be used in a way other than anticipated, with support workers helping the participants to complete five surveys (mental health, housing/homelessness, general health, alcohol/substance abuse, offending) to facilitate communication. Events that occurred and interventions performed by organisational staff were also recorded.

A secondary outcome from this trial was behavior change, which led to National Institute for Health and Care Excellence (NICE) recognition as an evaluation tool which facilitates behavioral change interventions [12]. Other benefits included participants feeling more listened to, and Lincus providing an auditable log of care provision.

The smooth running of this trial was facilitated by ongoing co-development and engagement. The existing culture of openness to technology and positive change played a major factor in this success. Communication also enabled co-developed adaptations to the tool to meet additional requirements.

Lessons learned include the potential for disconnect between the expectations of the technology developer and consumer [13]. Therefore, frequent engagement with stakeholders was vital throughout the technology design and implementation process.

2.2. Case Study 2

As an ongoing collaboration, Lincus was adapted and trialed with a learning disabilities charity to assess usability for individuals with learning disabilities living in supported accommodation [14]. The system was adapted through engagement with subject matter experts to develop survey questions and approach. Events and interventions were logged to enable their impact on wellbeing to be assessed. Support workers would assist end users to report information on Lincus.

Lincus demonstrated usability as a tool for communicating perceived overall health and wellbeing in this user group. As a result of the outcomes from this trial, Lincus was widely deployed by the charity.

During the trial many adaptations were made, and continue to be made, to improve usability. For example, no events or interventions were recorded during the first deployment. As a consequence, the system was developed to automatically prompt users after completing a survey to ask if they wished to add an event, which improved logging.

Following discussion with staff, many did not have the time to log in to monitor Lincus. Consequently, a new feature was developed to send reports from the system to the individual's email with desired information at a frequency that suited them. Considerable effort was made to improve end user engagement, including customisable interfaces, accessibility view, regular updates, tablet optimisation and new functionality.

Strong collaboration and flexibility in the design of the tool has resulted in long term commissioning of the technology, expanding use from a person centered recording platform to a total care management system.

2.3. Case Study 3

Lincus was further developed for individuals with long term conditions for use as part of a person centered coaching programme. Multiple changes were made to the system including interface redesign to make it more appealing to the wider population, and the integration of an activity tracker. The technology developers continued to adapt and modify the technology in line with feedback relating to user experience.

At completion of a limited 12 person trial, 100% of the participants said they would recommend the programme, stating that they had benefited from the intervention, and would continue to make improvements to their lifestyle based on what they had learnt. Integration of the activity tracker also led to self-reported behavioural change.

Though the whole coaching model has not been directly commissioned UK, the technology developments have been adopted by new and existing customers and partners.

2.4. Case Study 4

Lincus was tailored for a 12-month project with a City Council to support 300 individuals across a range of services, including supported living, care homes and young people during transition.

The project was commissioned by senior management without early inclusion of service providers who would be using the technology. This led to a lack of engagement at the start of the project due to service provider resource constraints. With further engagements, providers gained a better understanding of the system and how it could benefit them. It also changed the way the platform would be used for the project, with an increased focus on utilising the shared care capabilities, rather than self-care alone.

The authority also found it difficult to identify service providers, therefore the technology providers supported commissioners in identifying potential service providers where use could be beneficial.

After a revised project start date, multiple staff and service users have reported positive feedback when using the tool. However, outcomes of this project are yet to be identified.

Lessons learned include the need to involve those that will work with the system early on in the commissioning process to increase engagement during implementation. Organisations must also see the value in the tool, and must be willing to commit the required resource to adopt a new technology.

2.5. Case Study 5

Lincus has also been used as the underlying technology to deliver a Clinical Commissioning Group (CCG) sponsored programme to identify high blood pressure, and educate people on the importance of blood pressure in collaboration with an existing health service provider. Blood pressure screening clinics were set up around the city over a six-month period by the provider, and each person screened was given the opportunity to use Lincus. A tailored website was also developed with close consultation with the CCG and service provider to provide interactive information on blood pressure.

The programme was highly successful, with more than 1,000 people screened. Many individuals with high blood pressure were identified as a result of the screening. A key learning point from this programme is to work with a provider partner early on, and leverage existing services to co-develop a solution.

3. Discussion

There have been many lessons learned throughout the case studies outlined in this paper, which have informed development. The case studies highlight that each implementation is unique, therefore it is beneficial for the technology to be developed in a way that can be adapted to suit different services.

In order to achieve this, long term collaboration and engagement strategies must be utilised to develop technology that is designed to meet the needs of an organisation [9]. It is beneficial to engage with all stakeholders in this process, with the support of senior staff [4]. This can also help to overcome any concerns or issues, and identify more effective ways of using technology or integrating services early on in the implementation process [9]. Ongoing engagement has been a key reason for the success of a number of Lincus trials which have led to long term collaborations [12], [14].

Technology must also be adaptable to changing needs [9]. This can also lead to extended use of the technology within an organisation, as evidenced in Case Study 2.

Organisations need to be aware of the resource required to adopt a new technology, and be willing to dedicate internal resources to the implementation process [10]. Failure to do so can heavily impact implementation, especially in the early stages of adoption.

Finally, interoperability and integration with current systems, infrastructures, and ways of working are paramount for implementing new technologies in social and health care services [11]. It is vital adapt technology to fit workflows as closely as possible to ensure adoption, ongoing use, and benefit to organisations [13].

4. Conclusion

The case studies demonstrate that each implementation is different and unique. Collaborating closely with organisations and stakeholders on an ongoing basis is key to successful adoption. A shared learning process, and flexible system design allows adaptability for changing organisational needs. It is essential to design technologies to suit services and the complexity of their workflows and cultures to maximise success.

References

[1] WHO, Global action plan for the prevention and control of NCDs 2013-2020, 2013. [Online]. Available: http://www.who.int/nmh/publications/ncd-action-plan/en/

[2] Care Quality Commission, The state of health care and adult social care in England. 2015.

[3] Kings Fund, Technology in the NHS. Transforming the patient's experience of care. 2008. [Online]. Available: https://www.kingsfund.org.uk/sites/files/kf/Technology-in-the-NHS-Transforming-patients-experience-of-care-Liddell-Adshead-and-Burgess-Kings-Fund-October-2008_0.pdf

[4] D. A. Ludwick and J. Doucette, Adopting electronic medical records in primary care: lessons learned from health information systems implementation experience in seven countries, *Int. J. of Medical Informatics*, vol. 78, no. 1, pp. 22-31, Jan. 2009.

[5] M. B. Buntin et al., The benefits of health information technology: a review of the recent literature shows predominantly positive results, *Health Affairs*, vol. 30, no. 3, pp. 464-471, Mar. 2011.

[6] York Health Economics, Organisational and behavioural barriers to medical technology adoption, *NHS Institute for Innovation and Improvement*, pp. 1-144, Sep. 2009.

[7] The Kings Fund, A digital NHS? An introduction to the digital agenda and plans for implementation, 2016. [Online]. Available: https://www.kingsfund.org.uk/sites/files/kf/field/field_publication_file/A_digital_NHS_Kings_Fund_Sep_2016.pdf

[8] K. Or et al., Factors affecting home care patients' acceptance of a web-based interactive self-management technology, *J. of the Amer. Medical Informatics Assoc*., vol. 18, no. 1, pp. 51-59, Jan. 2011.

[9] K. Cresswell and A. Sheikh, Organizational issues in the implementation and adoption of health information technology innovations: an interpretative review, *Int. J. of Medical Informatics*, vol. 82, no. 5, pp. e73-e86, May 2013.

[10] D. Damschroder et al., Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science, *Implementation Science*, vol. 4, no. 1, Aug. 2009.

[11] Department of Health, Making IT work: harnessing the power of health information technology to improve care in England. 2016. [Online]. Available: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/550866/Wachter_Review_Accessible.pdf

[12] NICE, Lincus Trial: YMCA Liverpool pilot study for individuals with multiple and complex needs. 2013. [Online]. Available: https://www.nice.org.uk/sharedlearning/lincus-trial-ymca-liverpool-pilot-study-for-individuals-with-multiple-and-complex-needs

[13] K. Cresswell et al., Investigating and learning lessons from early experiences of implementing ePrescribing systems into NHS hospitals: a questionnaire study, *PLoS One*, vol. 8, no. 1, pp. e53369, Jan. 2013.

[14] L. Gilbert et al., Lincus: an adaptation and pilot usability study on individuals with learning disabilities, Rescon Internal Publication, Sep. 2014, unpublished.

This page intentionally left blank

2. Health Data Science

This page intentionally left blank

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-91

Predicting the Pathogenic Impact of Sequence Variation in the Human Genome

Mark F. ROGERS[a,1], Hashem A. SHIHAB[b], Michael FERLAINO[c,d], Tom R. GAUNT[b] and Colin CAMPBELL[a]
[a] *Intelligent Systems Laboratory, University of Bristol, Bristol, BS8 1UB, U.K.*
[b] *MRC Integrative Epidemiology Unit (IEU), University of Bristol, Bristol, BS8 2BN,U.K.*
[c] *Nuffield Department of Obstetrics and Gynaecology, John Radcliffe Hospital,University of Oxford, Oxford, OX3 9DU, U.K.*
[d] *Big Data Institute, University of Oxford, Oxford, U.K.*

Abstract. Sequencing data will become widely available in clinical practice within the near future. Uptake of sequence data is currently being stimulated within the UK through the government-funded 100,000 genomes project (Genomics England), with many similar initiatives being planned and supported internationally. The analysis of the large volumes of data derived from sequencing programmes poses a major challenge for data analysis. In this paper we outline progress we have made in the development of predictors for estimating the pathogenic impact of single nucleotide variants, indels and haploinsufficiency in the human genome. The accuracy of these methods is enhanced through the development of disease-specific predictors, trained on appropriate data, and used within a specific disease context. We outline current research on the development of disease-specific predictors, specifically in the context of cancer research.

Keywords. Prediction, sequence data, variant, annotation, point mutation, indel.

1. Introduction

Substantial improvements in sequencing technologies, and rapidly falling costs, will result in the widespread use of DNA sequence data within clinical practice. This trend is being encouraged within the UK through the Genomics England (100,000 genomes) project. Interpretation of these datasets poses challenges, from the size and complexity of the data through to the necessary linkage of DNA sequence data with other types of data, such as clinical covariates. For the analysis of DNA sequence data, a crucial challenge is the ability to distinguish which genetic variants are functional in disease, against a background of many disease-neutral variants. Accurate understanding of which genetic variants are pathogenic will improve our understanding of the molecular mechanisms underlying human disease and our ability to provide targeted therapies.

In recent research we have developed a variety of methods for predicting the pathogenic impact of genetic variants. In Shihab et al (2015) [1] we proposed an integrative classifier for predicting whether single nucleotide variants (SNVs) are

[1] Corresponding author, Intelligent Systems Laboratory, University of Bristol, Bristol, BS1 1UB, U.K.; E-mail: mark.rogers@bristol.ac.uk

functional in human disease, or neutral (for both coding or non-coding regions of the human genome). A number of sources of data are relevant to predicting if a SNV is functional in disease or is neutral. Consequently we used a variety of feature groups, or sources of data, which could be informative. In the construction of these prediction methods we used sequence conservation across species, histone modification (ChIP-Seq data), transcription factor binding site data, open chromatin data (DNase-Seq peak calls across cell lines from ENCODE), GC content, genome segmentation, and annotations describing DNA footprints across cell types (from ENCODE [3]). Thus, for example, sequence conservation across species proved to be a highly informative source: if a SNV occurs in a genomic region which is highly conserved across species it is much more likely to be functional in disease relative to a SNV which occurs in a region with high variability across species..

In our study, Shihab et al (2015) [1], we therefore used an algorithm-based approach capable of data integration i.e. the algorithm uses and learns to weight these different types of data, according to relative informativeness. In this study we used a specific approach to data integration called multiple kernel learning [2], though other data integration methods can be used. The method was called FATHMM-MKL (see fathmm.biocompute.org.uk for the prediction tool). Aside from giving a predicted label (pathogenic or neutral), the method also assigns a confidence measure to this label. At the default threshold on this confidence measure, FATHMM-MKL has a balanced test accuracy of 89.7%, with a false-positive rate of 3.8%. With a higher cutoff threshold on the confidence, the test accuracy slightly drops to 88.0% but with the false-positive rate dropping to 1.2%. A number of other groups have also proposed predictors for estimating the pathogenic impact of SNVs [4,5,6,7,8].

We have extended this line of investigation in a variety of directions. Small insertions and deletions (indels) can also have a significant influence in human genetic disease. In terms of relative frequency, indels are second only to SNVs as mutations. To date, classifiers for predicting the functional impact of indels have been restricted to their effect in the human exome (e.g. [9,10,11,12]). However, non-coding regions also contain many functional elements. Indeed, the vast majority of catalogued SNV-trait associations fall within non-coding regions of the human genome [13]. We have proposed an integrative predictor for estimating the pathogenic impact of indels in non-coding regions of the human genome [14]: the method is called FATHMM-indel and is available via the Web (indels.biocompute.org.uk). Using nested cross validation, this classifier achieves a balanced accuracy of 86%. In other work [15] we have proposed a Genome Tolerance Browser to visualise the possible pathogenic impact of SNVs in the genome (this tool is available at gtb.biocompute.org.uk). A further project has been to develop a state-of-the-art predictor, called HiPred, for estimating the effect of haploinsufficiency [16]. Cells in the human body are diploid, they contain two complete sets of chromosomes, one from each parent. Haploinsufficiency occurs if there is only one functional copy of a gene, and this single copy does not produce a sufficient amount of a gene product, resulting in a disease trait.

2. Disease-specific prediction

The predictors for SNVs and indels have a high accuracy in many simple disease contexts but are still not sufficiently accurate when applied to more complex multifactorial diseases. For a complex disease, such as cancer, oncogenesis is typically driven by a

combination of disease-enabling genetic variants. For construction of a prediction tool, this creates a label-dependency problem during classifier training: a single point mutation could be labelled pathogenic or neutral, depending on the labels at other locations in the cancer genome. In any case, training a classifier with domain-specific data, such as sequence data exclusively from a particular type of cancer, would likely offer improved accuracy. Indeed, our previous studies have suggested that disease-specific predictors are more accurate than generic predictors [18].

With this motivation we are devising cancer-specific predictors, for predicting the oncogenic impact of single point mutations. As for FATHMM-MKL, these predictors are trained using a variety of data sources falling into three main categories: genomic (genomic features include GC content, sequence spectra, repeat regions and measures of region uniqueness), evolutionary (as for FATHMM-MKL, evolutionary features include a comprehensive set of conservation-based measures) and consequences (for coding regions only: data derived from the Variant Effect Predictor [19] and other sources). To train the classifier for handling single point somatic mutations, we used high recurrence rate SNVs from the COSMIC cancer database [20] as the positives, with negatives derived from the 1000 Genomes project [21]. We achieved state-of-the-art performance with a substantial gain over competitors (Figure 1, left). This predictor is called CScape and is available via the Web (cscape.biocompute.org.uk) [22]. Evaluated via leave-one-chromosome-out cross-validation (LOCO-CV), the approximately balanced test accuracy is 72.3% in coding regions and 62.9% in non-coding regions. As with FATHMM-MKL we also devised a confidence measure associated with the predicted class label.

Though promising, the test accuracy of the resultant classifiers remains inadequate for use by cancer researchers. However, if we restrict prediction to the highest confidence instances then it is possible to achieve 91.7% test accuracy (with LOCO-CV, coding regions only). Given a positive predictive value (PPV) of 0.78, and a large number of true positives, this test accuracy is not achieved by predominant accurate prediction of negatives (non-oncogenic single point mutations). This strong performance comes at the expense of yielding predictions for just 17.7% of coding region nucleotide positions (Figure 1, right). Nevertheless, this becomes an experimentally usable level of accuracy.

However, this classifier (cscape.biocompute.org.uk) still remains generic, in that it is trained on COSMIC data [20], derived from a variety of cancer types. Thus further improvement can be achieved by developing predictors trained on, and specific to, individual types of cancer. As an example, using data from the Cancer Genome Atlas [22] and the International Cancer Genome Consortium [23] we have derived specialist predictors for particular types of cancer. Thus for a specialised breast cancer predictor (CScape-brca), we can achieve a baseline predictor (coding regions, all nucleotide positions) with approximate 80% accuracy and capable of a greater test accuracy, if restricted to higher confidence predictions.

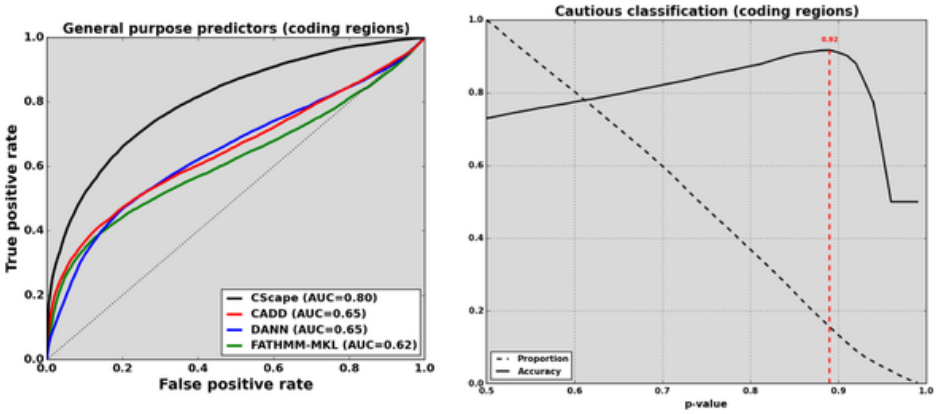


Figure 1. Left: ROC curves for a comparison of the proposed classifier (*CScape*) for predictions in the coding regions of the cancer genome, against alternative methods. Right: the solid curve gives the test accuracy (approximately balanced), the dashed curve gives the proportion of nucleotide positions with high enough confidence for prediction at the given level of test accuracy: this dashed curve is derived from test data, the 12.1% figure quoted in the text is for whole-genome prediction (coding regions).

3. Discussion

These new methods indicate that usable levels of accuracy can be achieved for predicting the pathogenic impact of genetic variants. Aside from predicting possible new drug targets, the refined insights from these tools could assist in establishing subtypes of disease, hence improving personalised approaches to therapy and predicting an individual's response to a drug. There is the prospect that these methods can be further enhanced through the incorporation of additional sources of data. Aside from disease-specific prediction, another avenue for investigation would be region-specific prediction, for example, dedicated predictors for non-coding variants residing at or near splicing regions. We will report on these developments in later work.

References

[1] H. Shihab, M. Rogers, J. Gough, M. Mort, D. Cooper, I. Day, T. Gaunt and C. Campbell, An integrative approach to predicting the functional effects of non-coding and coding sequence variation, *Bioinformatics*, **31** (2015), 1536-1543.
[2] C. Campbell, C. and Y. Ying, *Learning with Support Vector Machines*, Morgan and Claypool, 2011.
[3] The ENCODE Project Consortium, An integrated encyclopedia of DNA elements in the human genome, *Nature*, **489** (2012), 57–74.
[4] I.A. Adzhubei, S. Schmidt, L. Peshkin, V.E. Ramensky, A. Gerasimova, P. Bork, A.S. Kondrashov and S.R. Sunyaev, S.R., A method and server for predicting damaging missense mutations, *Nature Methods*, **7** (2010), 248-249.
[5] P. Kumar, S. Henikoff and P.C. Ng, Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm, *Nature Protocols*, **4** (2009), 1073–81.
[6] B. Reva, Y. Antipin and C. Sander, Predicting the functional impact of protein mutations: application to cancer genomics, *Nucleic Acids Research*, **39** (2011), e118.
[7] M. Kircher, D. Witten, P. Jain, B. O'roak, G. Cooper, G and J. Shendure, A general framework for estimating the relative pathogenicity of human genetic variants. *Nature genetics*, **46**, (2014) 310-315.
[8] D. Quang, D., Y. Chen X. and Xie. DANN: a deep learning approach for annotating the pathogenicity of genetic variants, *Bioinformatics*, **31** (2014), 761–763.

[9] Y. Choi and A.P. Chan, PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels, *Bioinformatics*, **31** (2015), 2745–2747.
[10] C. Douville, D.L. Masica, P.D. Stenson, D.N. Cooper, D.G. Gygax, R. Kim, M. Ryan and R. Karchin, Assessing the Pathogenicity of Insertion and Deletion Variants with the Variant Effect Scoring Tool (VEST-indel), *Human Mutation*, **37** (2016), 28–35.
[11] Folkman, Y. Yang, Z. Li, B. Stantic, A. Sattar, M. Mort, D.N. Cooper, Y. Liu and Y. Zhou, DDIG-in: detecting disease-causing genetic variations due to frameshifting indels and nonsense mutations employing sequence and structural properties at nucleotide and protein levels, *Bioinformatics*, **31** (2015), 1599–1606.
[12] H. Zhao, Y. Yang, H. Lin, X. Zhang, M. Mort, D.N. Cooper, Y. Liu and Y. Zhou, DDIG-in: discriminating between disease-associated and neutral non-frameshifting micro-indels, *Genome Biology*, **14** (2013), R23.
[13] L. Hindorff, P. Sethupthay, H. Junkins, E. Ramos, J. Mehta, F. Collins and T. Manolio, Potential etiologic and functional implications of genome-wide association loci for human diseases and traits, *PNAS*. U.S.A., **106** (2009), 9362–9367.
[14] M. Ferlaino, M. Rogers, H. Shihab, T. Gaunt, M. Mort, D. Cooper and C. Campbell. An integrative approach to predicting the functional effects of small indels in non-coding regions of the human genome. *Journal submission* (2017).
[15] H. Shihab, M. Rogers, C. Campbell and T. Gaunt. GTB - An Online Genome Tolerance Browser. *BMC Bioinformatics* **18** (2017), 20.
[16] H. Shihab, M. Rogers, C. Campbell and T. Gaunt, HiPred: an integrative approach for predicting haploinsufficiency in the human genome. *Bioinformatics*, doi.10.1093/bioinformatics (2017).
[17] M. Rogers, H. Shihab, T. Gaunt and C. Campbell, CScape: a tool for predicting oncogenic single-point mutations in the cancer genome. *Journal submission* (2017).
[18] H. Shihab, J. Gough, M. Mort, D. Cooper, I. Day and T. Gaunt, Ranking Non-Synonymous Single Nucleotide Polymorphisms based on Disease Concepts, *Human Genomics*, **8** (2013), 11.
[19] http://www.ensemb.org/info/docs/tools/vep/index.html
[20] http://cancer.sanger.ac.uk/cosmic/help/gene/analysis
[21] The 1000 Genomes Project Consortium, An integrated map of genetic variation from 1,092 human genomes, *Nature*, **491** (2012), 56–65.
[22] https://cancergenome.nih.gov/
[23] http://icgc.org/

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-96

Learning Healthcare System for the Prescription of Genetic Testing in the Gynecological Cancer Risk

Cristina SUÁREZ-MEJÍAS[a], Alicia MARTÍNEZ-GARCÍA[a,1], María Ángeles MARTÍNEZ-MAESTRE[b], José Manuel SILVAN-ALFARO[b] , Jesús MORENO CONDE[a] and Carlos Luis PARRA-CALDERÓN[a]
[a] *Group of Research and Innovation in Biomedical informatics, biomedical engineer and health economy. Institute of Biomedicine of Seville, IBiS/"Virgen del Rocío" University Hospital /CSIC/University of Seville. Seville, Spain*
[b] *Gynecology Department, Virgen del Rocío University Hospital, Seville, Spain*

Abstract. Clinical evidence demonstrates that BRCA 1 and BRCA2 mutations can develop a gynecological cancer but genetic testing has a high cost to the healthcare system. Besides, several studies in the literature indicate that performing these genetic tests to the population is not cost-efficient. Currently, our physicians do not have a system to provide them the support for prescribing genetic tests. A Decision Support System for prescribing these genetic tests in BRCA1 and BRCA2 and preventing gynecological cancer risks has been designed, developed and deployed in the Virgen del Rocío University Hospital (VRUH). The technological architecture integrates a set of open source tools like Mirth Connect, OpenClinica, OpenCDS, and tranSMART in addition to several interoperability standards. The system allows general practitioners and gynecologists to classify patients as low risk (they do not require a specific treatment) or high risk (they should be attended by the Genetic Council). On the other hand, by means of this system we are also able to standardize criteria among professionals to prescribe these genetic tests. Finally, this system will also contribute to improve the assistance for this kind of patients.

Keywords. Learning Healthcare System, genetic testing, gynecological cancer, breast cancer, ovarian cancer

1. Introduction

Breast Cancer (BC) is the most common tumor which affects to women in the western world. Ovarian Cancer (OC) is the fifth malign tumor in women. However, difficulties in its diagnosis and its therapeutic treatment imply a high mortality, greater than 50% at 5 years since diagnosis. The International Agency for Research on Cancer estimated a worldwide incidence of 1.67 million new BC cases diagnosed per year and over 0.23 million of OC [1]. Although the appearance of these tumors is usually sporadic, around 10-15% of diagnosed cases are heritable. BRCA1 and BRCA2 genes described germinal mutations which are inhered dominantly and with a high penetrance in 7% of BC and about in 11-15% of OC [2]. In fact, BRCA1 and BRCA2 mutations increase

[1] Corresponding author, e-mail: alicia.martinez.exts@juntadeandalucia.es.

the risk of BC and/or OC. Concretely; people with BRCA1 mutations have a 57% risk of developing BC and around 40% of developing OC. For BRCA2 mutations, the risk is 49% in BC and of 18% in OC [3-5]. Another significant element is its behavior pattern. Hereditary gynecologic cancer usually starts in younger people while this trend is not usual for sporadic cancer. Also, it has a more invasive histopathological pattern. However, its diagnosis is important because the answer could be positive with determined treatments [6]. Due to it, it is important for people with family history to know the risk of having a hereditary cancer. Moreover, if the risk is not known, it will provoke patient anxiety and concern. Each treatment of patients is different, because it depends on the BRCA1 and BRCA2 gene mutation. If the mutation is positive, patients will receive monitoring or preventative measures. If the result is negative, it will decrease the anxiety level of patient. In this sense, there are some studies which demonstrated that patients are benefited by the genetic testing results [7-9]. These studies found out a significant decrease in patient concern about developing cancer. In clinical practice, general practitioners and gynecologists are the first in the attendance of patient concern. However, these genetic tests have a high cost to the healthcare system. Besides, several studies in the literature indicate that performing these genetic tests to the population is not cost-efficient [10]. In this sense, criteria were defined to identify patients with high risk for developing these mutations. For this purpose, there were consensus between the scientific society and the official organization in the definition of these criteria. [11]. Genetic testing is only recommended for patients with previous family and personal history. Providing to physicians a Learning Healthcare System (LHS) to improve the performance of genetic testing is the key to optimize the prescription of this kind of tests. Authors of this paper have experience in the development of a LHS for pulmonary thromboembolism in a context of intrahospital use [13].

2. Method

In this section the implemented LHS to support physicians about medical prescription of genetic testing is described. Our proposal is based on an open source Clinical Decisions Support System (CDSS), a service-oriented architecture and the use of interoperability standards defined by HL7 (Virtual Medical Record (vMR)) and OMG [14]. These standards are being materialized through the initiative OpenCDS [15]. Besides, the set of data registered and decision rules are defined. The CDSS defines the level of risk of possible existence of BRCA1 or BRCA2 gene mutation. In addition, it will generate an information document for the evaluated patient. Those patients with low risk will receive a document with general recommendations whicht is identical to information provided to healthy population. On the other hand, those patients with high level of risk will receive information and the possibility to be derived to the hereditary familial Gynecological Pathology Consultation for proper genetic counseling and request of genetic testing. Furthermore, the LHS includes tranSMART [16] for analyzing data and improving the decision rules based on knowledge.

2.1. Architecture of LHS

In the technological architecture (Figure 1), some open source tools were integrated: 1) Mirth. An enterprise service bus (ESB) which manages the communication and integration between the different components of the whole architecture; 2) OpenClinica. A clinical trial management system that implements the end user interface to allow physicians register patient information through electronic forms; 3) OpenCDS. A CDSS that executes rules modeled by JBoss Guvnor and following the vMR reference model. This system executes the decision support rules taking into account not only the information registered by means of OpenClinica but also the information extracted from the Hospital Systems.; and 4) JBoss Drools Guvnor. A rule modeler that imports the vMR reference model to design decision support rules.. Besides, a custom database has also been used to map data from OpenClinica to vMR format. Both clinical and technical teams have worked together to model in Guvnor the decision support rules related to gynecological cancer. All these rules were automatically accessible through OpenCDS. Finally, the LHS also integrates tranSMART, a translational biomedical research datawarehouse that integrates data from heterogeneous data sources in order to provide end users the capabilities to search, view, and analyze all data stored within.

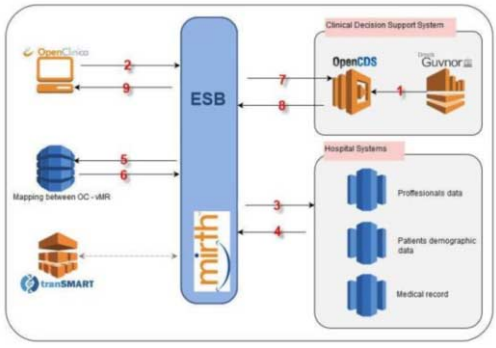


Figure 1. LHS architecture.

2.2. Decision Rules Implemented

The minimum set of data and decision support rules used in the CDSS were designed taking into account the guideline defined by Spanish Society of Medical Oncology [17]. This guideline evaluates patient risks and recommends the performance of a genetic test. The system also records other genetic and clinical information about patients. The hypothesis establishes that data mining analytical techniques could evolve, based on evidence, the minimal set of data and rules required. It could also ensure the continuous improvement of effectiveness of the developed system. According to the guideline of the SEOM, the information recorded in the system is the following one:

- Patient with BC or OC: Woman diagnosed of papillary serous OC of high grade; or woman diagnosed of BC before 30 years old; or woman diagnosed of BC and OC; or woman diagnosed of bilateral BC, when one of the tumor was before 40 years old; or woman diagnosed of triple negative BC before 50 years old; or men diagnosed of BC.

- Undiagnosed patient of cancer. Woman with 2 cases of cancer in first degree relatives and also: one of the tumors was diagnosed before 50 years old; or there was one relative with BC and OC; or there was one male relative with BC and there was another relative with BC or OC; or there were 3 or more relatives with BC where at least 2 of them were first degree relatives.

The CDSS will recommend performing a genetic test if the patient accomplishes any of before criteria. On the other hand, other information is also registered in the LHS for the generation of new hypothesis: age of diagnosis in case the patient has BC or OC, first and second degree relatives with BRCA1 or BRCA2 mutations (name and type of the mutation using Human Genome Variation Society) and relatives with any kind of tumors (type, age of diagnosis and kinship). All this information is also integrated in tranSMART, which provides users the capability to analyze all the information and generate new knowledge for improving in a continuous way the LHS.

3. Results

The LHS has been designed, implemented and deployed in the VRUH infrastructure. Currently, this LHS is accessible for our physicians and we plan to start the pilot stage in the following months. During this pilot, we are going to perform the following flow (Figure 1). 1) The specific rules modeled with Guvnor are automatically available to use by OpenCDS; 2) The healthcare professionals record information in the defined forms; 3) The ESB will connect with the hospital systems to get additional patient data; 4) The ESB extract the information from the hospital systems; 5) There is executed an automated process to map information from OpenClinica format (including both data that the healthcare professional has completed in the OpenClinica web interface, and data extracted from the hospital systems) to vMR format; 6) The ESB generate and XML structure in vMR format; 7) The ESB calls the OpenCDS services; 8) OpenCDS provides a recommendation from the CDSS; 9) The clinical decision support recommendation is shown to the healthcare professional. Finally, physicians analyze data and propose new hypothesis through tranSMART tool.

4. Discussion

In order to use OpenCDS, it is necessary to model the decision support rules by means of a HL7 reference model known as vMR. We must realize that vMR is very extensive due to it try to cover all the necessary concepts and entities in a health scenario. As a result, vMR is presented in an abstract way that, although it is very interesting in a conceptual point of view, it might be difficult to manage. As a consequence, the rule modeling process for a real scenario using vMR and a mapping process between OpenClinica and vMR format have been very hard tasks. In fact, the new version of OpenCDS is migrating from vMR to HL7 Fast Healthcare Interoperability Resources [18] reference model, a new standard easy to learn, adapt and implement.

5. Conclusion

A LHS for prescribing BRCA1 and BRCA2 genetic testing has been developed. This system allows general practitioners and gynecologists to decide about performing these genetic tests in the prevention phase and early detection of BC and OC. For this purpose, the system was developed using OpenCDS and others open source tools. Nowadays, the LHS is being piloted with real cases. It will permit to standardise the criteria and improve the process of assistance of these patients. We have also surveyed final users through a Technology Acceptance Model to evaluate the opinion of the LHS.

Acknowledgements

The research was supported by grants to from the Spanish Ministry of Economy and Competitiveness, Plan Estatal de I+D+I 2008-2011, ISCIII-AES 2012 PITeS ISA: definición, diseño y desarrollo de herramientas y servicios basados en estándares para el apoyo a la decisión clínica y medicina personalizada project (PI12/01571) project coordinator Carlos Parra and Plan Estatal de I+D+I 2013-2016, ISCIII-AES 2013 ITEMAS ITEMAS: Platform for Innovation in Medical Technologies and Health (PT13/0006/0036) project coordinator Sandra Leal, both co-funded by FEDER from European Regional Development Funds (European Union).

References

[1] GLOBOCAN 2012, International Agency for Research on Cancer http://globocan.iarc.fr/
[2] G. Llort, I. Chirivella, R. Morales, SEOM clinical guidelines in Hereditary Breast and ovarian cancer, *Clin Transl Oncol*, **17**:12 (2015), 956-61.
[3] T.R. Rebbeck, N. Mitra, Wan F et al, Association of type and location of BRCA1 and BRCA2 mutations with risk of breast and ovarian cancer, *JAMA* **313**:13 (2015), 1347-61.
[4] S.A. Narod, Modifiers of risk of hereditary breast cancer, *Oncogene* **25**(2006), 5832-6.
[5] N. Andrieu, D.E. Goldgar, D.F. Easton et al. Pregnancies, breast-feeding, and breast cancer risk in the International BRCA1/2 Carrier Cohort Study (IBCCS), *J Natl Cancer Inst*, **98**:8 (2006), 535-544.
[6] J.L. Profato, B.K. Arun, Genetic risk assessment for breast and gynecological malignancies, *Curr Opin Obstet Gynecol* **27**:1 (2015), 1-5.
[7] S. Pruthi, B.S. Gostout, N.M. Lindor, Identification and Management of Women With BRCA Mutations or Hereditary Predisposition for Breast and Ovarian Cancer, *Mayo Clin Proc* **85**:12 (2010), 1111-20.
[8] C Foster, D.G. Evans, R. Eeles,D. et al, Predictive testing for BRCA1/2: attributes, risk perception and management in a multi-centre clinical cohort, *British Journal of Cancer*, **8** (2002), 1209-1216.
[9] M. Watson, C. Foster, et al Psychosocial impact of breast/ovarian (BRCA1/2) cancer-predictive genetic testing in a UK multi-centre clinical cohort, *British Journal of Cancer* **91** (2004), 1787-94.
[10] M.J. Esplen, J. Hunter, et al, A multicenter study of supportive- expressive group therapy for women with BRCA1/BRCA2 mutations, *Cancer* **101**:10 (2004), 2327-40.
[11] I. Ejarque, M. García-Ribes, J.V. Sorli, E. Arenas, V. Martín, El papel de la atención primaria ante el cáncer hereditario, *Atención Primaria* **40** (2008), 525-9.
[12] A. García-Jimenez et al, Clinical Decision Support using a Terminology Server to improve Patient Safety, EFMI, MIE 2015, IOS Press, 2015.
[13] HL7 International, *Clinical Decision Support; vMR Logical Model, Release 2*, January 2014.
[14] Kawamoto, K., et al. Evaluation of HL7 Continuity of Care Document as the Foundation for an International Virtual Medical Record Standard for Clinical Decision Support.
[15] Transmart http://transmartfoundation.org/
[16] Spanish Society of Medical Oncology (SEOM) http://www.seom.org/
[17] HL7 International, HL7 Fast Healthcare Interoperability Resources Specification (FHIR), Release 1, February 2016.

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-101

Exploratory Clustering for Patient Subpopulation Discovery

Dragan GAMBERGER[a,1], Bernard ŽENKO[b], Nada LAVRAČ[b,c], and for the Alzheimer's Disease Neuroimaging Initiative[2]
[a] *Rudjer Bošković Institute, Croatia*
[b] *Jožef Stefan Institute, Slovenia*
[c] *University of Nova Gorica, Slovenia*

Abstract. Exploratory Clustering is a novel general purpose clustering tool which is especially appropriate for medical domains in which we need to identify subpopulations that are similar in two different data layers. The tool implements the multi-layer clustering algorithm in a framework that enables iterative experiments by the user in his search for relevant patient subpopulations. A unique property of the tool is integration of clustering and feature selection algorithms. Differences in values of most relevant attributes are used to demonstrate decisive properties of constructed clusters. Usefulness of the tool is illustrated on a task of discovering groups of patients with similar cognitive impairment.

Keywords. Data clustering, biomarkers, Alzheimer's disease

1. Introduction

In this work we present a novel publicly available web application for data clustering, which is useful for detection of relevant subpopulations that are similar in two different data layers at the same time. A typical application domain is medicine where, for example the first layer comprises biological or genetic data while the second layer comprises clinical data. Detection of subpopulations homogeneous in these two layers is relevant for understanding relations between biological and clinical variables and for biomarker identification. If the objective of data analysis is medical prognosis, then the first layer can consist of baseline patient information while the second layer can contain corresponding longitudinal data. A nice property of this approach is that if the resulting clusters are homogeneous at the same time in different data layers, the quality of clustering increases (e.g., in multi-view clustering [1] and redescription mining [2]).

It is known that objective evaluation of the quality of clustering is practically impossible [3]. For the same data different solutions are possible and selection of the optimal one depends on human understanding of the data analysis problem, meaning that

[1] Corresponding author: Dragan Gamberger, Rudjer Bošković Institute, Bijenička 54, 10 000 Zagreb, Croatia; E-mail: dragan.gamberger@irb.hr.

[2] Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: `http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf`

human expert knowledge is essential for high quality clustering. Our goal when developing the Exploratory Clustering tool was to design an extremely simple tool that medical researchers will be able to use by themselves. This should make it easier to generate medically and scientifically relevant data analysis results.

Section 2 presents the basic concepts underlying the implemented tool, Section 3 describes the data upload page, while Section 4 presents and discusses the results for a small set of patients with cognitive impairment extracted from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database [4].

2. System Description

Exploratory Clustering is a web application, therefore the user does not need to download and install any software. Instead he uses the web browser to upload data to the computing server, to interactively guide the analysis process and to get the results of the analysis.

The analysis with the Exploratory Clustering tool is an iterative process. The user uploads data and receives a result that is optimal according to the implemented clustering algorithm. In the next step the user can ask for the refinements of the current solution. The refinements can go in two directions. Either the user can ask for modification of the current solution by increasing or decreasing the size of the constructed clusters or he can ask for a new clustering solution from a different subset of input data. The process can be iterated many times, enabling the user to employ his expert preferences in order to select the optimal clustering result from a large set of potentially good solutions. It must be noted that the user selects the direction in which the refinements should be executed, while the clustering algorithm determines how each refinement is actually implemented. This ensures that results of all iterations reflect relations existing in the data and present potentially good solutions.

Exploratory Clustering combines clustering and feature selection algorithms. Integration of feature selection into the clustering process is important because it enables detection and elimination of irrelevant variables, making it possible to cluster also high dimensional data where instances are described by many variables (attributes). Additionally, this approach enables detection of variables that are most responsible for the current clustering result. By showing these variables to the user and especially by computing and presenting their average values (or mode values for categorical variables) for each cluster, the user can better understand the meaning of the constructed clusters and significance of differences among them. Specifically for exploratory clustering this information is of ultimate relevance for the user because it is the basis for selecting the optimal solution.

The tool is based on the multi-layer clustering methodology described in [5, 6]. We decided to use this methodology because it enables both single and two layer clustering and because it can work with correlated layers (e.g., in multi-view clustering correlations between views are not allowed [1]). The second property is important especially for medical applications. In contrast to most other clustering tools [7], the multi-layer clustering algorithm determines the number of clusters and their optimal size automatically, thus users do not have to adjust any parameters of the clustering algorithm. In the final result some or even many instances may remain unclustered. In this way the constructed clusters correspond to sets of similar instances, while other instances remain unclustered. In some cases unclustered instances may be interpreted as outliers.

3. Exploratory Clustering Web Application

Exploratory Clustering tool is available at `http://rr.irb.hr/exploC/`. Because of the limited space we are not able to include the screen-shot of the data upload page but the reader can check it on the web. The page also has the link to instructions for data preparation, which include two tutorials describing the tool and its application.

In its basic form the Exploratory Clustering can be used as a standard clustering tool for data sets with up to 1,000 instances and up to 1,000 attributes. In this case it is only necessary to specify a data file for layer 1. Optionally, the user can upload also a file with the names of attributes, a file with the names of instances, and a file with some known classification of examples. Upload of optional files does not affect the clustering result but it can increase the understandability of the results that are presented to the user.

For two-layer clustering the user has to prepare and upload also the data file for layer 2. If biological data are uploaded in layer 1 then layer 2 is for clinical data or if baseline data are in layer 1 then longitudinal data are in layer 2. The second layer can include also up to 1,000 attributes. Optionally, the user can upload also the names of attributes in the second data layer.

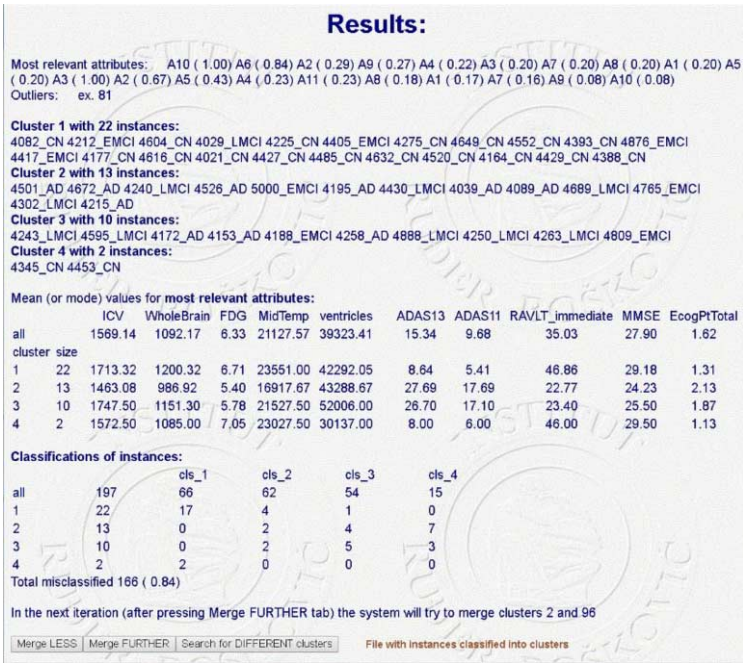
The user does not have to specify any parameters but can select increased reliability of the results. Increased reliability means execution of more iterations for computation of the similarity of instances [5, 6]. With this option the computation takes more time and its use is not recommended for data sets with more than 500 instances.

4. Illustrative Example

A data set of 197 male patients that have problems with dementia in used to illustrate the use of the tool. The data set is a subset of patients from the ADNI database [4] for which extensive clustering experiments have been performed and already reported in [5, 6]. In the first layer are 15 biological measurements like ABETA peptides, TAU and PTAU proteins, and MRI volumetric data together with 41 laboratory variables like number of red blood cells and total bilirubin values. In the second layer are 147 clinical variables like Alzheimer's Disease Assessment Scale (ADAS13) and Mini Mental State Examination (MMSE) score together with 40 symptoms like nausea and dizziness.

Besides biological and clinical data we also upload attribute names for both layers, names of examples and classification of examples according to the medical diagnosis that is not used as input data for clustering. Names of examples are a combination of the patient's RID number and the medical diagnosis that can be CN (cognitive normal), EMCI (early mild cognitive impairment), LMCI (late mild cognitive impairment) or AD (Alzheimer's disease). Classification of instances is in four classes so that patients with diagnosis CN are in class 1 while AD patients are class 4.

Figure 1 illustrates clustering results obtained on the described data set. The central part of the report is the list of constructed clusters. Each cluster is represented by a list of included instances. In this case the solution consists of four clusters with a total of 47 instances. The result demonstrates a high non-homogeneity of input instances with 150 out of 197 instances remaining unclustered. If the user is not satisfied by such weak clustering result he can iteratively press the tab "Merge FURTHER" at the bottom of the web page. In this way he can get even a solution with all 197 instances in only 2 clusters.



Results:

Most relevant attributes: A10 (1.00) A6 (0.84) A2 (0.29) A9 (0.27) A4 (0.22) A3 (0.20) A7 (0.20) A8 (0.20) A1 (0.20) A5 (0.20) A3 (1.00) A2 (0.67) A5 (0.43) A4 (0.23) A11 (0.23) A8 (0.18) A1 (0.17) A7 (0.16) A9 (0.08) A10 (0.08)
Outliers: ex. 81

Cluster 1 with 22 instances:
4082_CN 4212_EMCI 4604_CN 4029_LMCI 4225_CN 4405_EMCI 4275_CN 4649_CN 4552_CN 4393_CN 4876_EMCI 4417_EMCI 4177_CN 4616_CN 4021_CN 4427_CN 4485_CN 4632_CN 4520_CN 4164_CN 4429_CN 4388_CN
Cluster 2 with 13 instances:
4501_AD 4672_AD 4240_LMCI 4526_AD 5000_EMCI 4195_AD 4430_LMCI 4039_AD 4089_AD 4689_LMCI 4765_EMCI 4302_LMCI 4215_AD
Cluster 3 with 10 instances:
4243_LMCI 4595_LMCI 4172_AD 4153_AD 4188_EMCI 4258_AD 4888_LMCI 4250_LMCI 4263_LMCI 4809_EMCI
Cluster 4 with 2 instances:
4345_CN 4453_CN

Mean (or mode) values for **most relevant attributes:**

		ICV	WholeBrain	FDG	MidTemp	ventricles	ADAS13	ADAS11	RAVLT_immediate	MMSE	EcogPtTotal
all		1569.14	1092.17	6.33	21127.57	39323.41	15.34	9.68	35.03	27.90	1.62
cluster	size										
1	22	1713.32	1200.32	6.71	23551.00	42292.05	8.64	5.41	46.86	29.18	1.31
2	13	1463.08	986.92	5.40	16917.67	43288.67	27.69	17.69	22.77	24.23	2.13
3	10	1747.50	1151.30	5.78	21527.50	52006.00	26.70	17.10	23.40	25.50	1.87
4	2	1572.50	1085.00	7.05	23027.50	30137.00	8.00	6.00	46.00	29.50	1.13

Classifications of instances:

		cls_1	cls_2	cls_3	cls_4
all	197	66	62	54	15
1	22	17	4	1	0
2	13	0	2	4	7
3	10	0	2	5	3
4	2	2	0	0	0

Total misclassified 166 (0.84)

In the next iteration (after pressing Merge FURTHER tab) the system will try to merge clusters 2 and 96

Merge LESS | Merge FURTHER | Search for DIFFERENT clusters | File with instances classified into clusters

Figure 1. Clustering result for 197 male patients with cognitive problems.

By checking the names of instances included into clusters we can conclude that constructed clusters are pretty consistent in respect of the diagnosis. In clusters 1 and 4 are mostly CN patients, in cluster 2 are mostly AD patients while in cluster 3 are mostly LMCI patients. Because we have prepared the file in which different diagnoses are coded with values 1-4 we have enabled generation of the classification report at the bottom of the web page. From this report it is easy to assess the consistency of clusters. For example, we see that in the largest cluster with 22 instances there are 17 CN patients, 4 EMCI patients and 1 LMCI patient. There are 166 misclassified instances, which corresponds to the sum of the number of unclustered instances and the number of minority class instances in all the clusters.

For expert evaluation the most interesting part of the report is the list of 5 most relevant attributes from each layer. For these 10 attributes the tool computes their average values for all 197 instances and then its average value for every constructed cluster. Large differences between reported values mean that the tool has been successful in detecting clusters that are substantially different. For example, for attribute ADAS13 the average value for all instances is 15.34 while average values for clusters 1 and 4 are about 8 and for clusters 2 and 3 the average values are about 27. But the data may reveal also some unexpected properties of constructed clusters. For example, for attribute ICV we have the average value for all instances 1,569, for cluster 2 with majority of AD patients we have substantially lower value 1,463 while for cluster 3 with majority of LMCI patients we have a substantially increased value equal to 1,747. In contrast, for attribute ventricles all clusters 1-3 have values higher than the average value for the complete population with highest value being 52,006 for cluster 3. This information can be very interesting for expert evaluation and for the user's decision if the constructed clusters are relevant.

5. Conclusion

To the best of our knowledge the Exploratory Clustering is the only clustering tool available as a web application, the tool that besides clusters of instances themselves also presents characterization of the constructed clusters, and the only tool that enables effective search for optimal solution over a set of different potentially good solutions. A simple user interface and parameter free clustering algorithm are additional advantages of the tool. Integration of feature selection into the clustering algorithm enables that in contrast to many other clustering algorithms that have a problem with the curse of dimensionality this tool can be used also for data sets with a large number of non-informative variables.

A serious drawback is time complexity of the tool, which is growing fast with the number of instances. An additional problem, especially when the data set has many variables, is that the refined solutions can only be slight modifications of the current solution and the user has to go through many iterations in order to get substantially novel clusters.

Acknowledgements

The authors acknowledge the financial support from the Slovenian Research Agency core research programme *Knowledge Technologies* (P2-0103) and project *HinLife: Analysis of Heterogeneous Information Networks for Knowledge Discovery in Life Sciences* (J7-7303); the European Commission's support through *The Human Brain* project (FET Flagship grant FP7-ICT-604102), *MAESTRA* project (Gr. no. 612944), and *InnoMol* project (Gr. no. 316289); support of the Croatian Science Foundation project *Machine Learning Algorithms for Insightful Analysis of Complex Data Structures* (Gr. no. 9623).

References

[1] S. Sun, A survey of multi-view machine learning, *Neural Computing and Applications* **23** (2013), 2031-2038.

[2] L. Parada and N. Ramakrishnan, Redescription mining: structure theory and algorithms, *Proceedings of the association for the advancement of artificial intelligence* AAAI '05 (2005), 837–844.

[3] U. von Luxburg, R.C. Williamson and I. Guyon, Clustering: Science or art?, *In Guyon, I., Dror, G., Lemaire, V., Taylor, G. W., and Silver, D. L. (eds.), ICML Unsupervised and Transfer Learning* **27** (2012), 65–79.

[4] M.W. Weiner et al., The Alzheimer's Disease Neuroimaging Initiative: a review of papers published since its inception, *Alzheimer's Dementia* **8** (2012), S1-68.

[5] D. Gamberger, B. Ženko, A. Mitelpunkt and N. Lavrač, Homogeneous clusters of Alzheimer's disease patient population, *Biomedical Engineering Online* **15** (2016) S78.

[6] D. Gamberger, B. Ženko, A. Mitelpunkt, N. Shachar and N. Lavrač, Clusters of male and female Alzheimer's disease patients in the Alzheimer's Disease Neuroimaging Initiative (ADNI) database, *Brain Informatics* **3** (2016), 169-179.

[7] G. Gan, C. Ma and J. Wu, *Data Clustering: Theory, Algorithms, and Applications*, SIAM Philadelphia, 2007.

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-106

An Automatic Approach for Analyzing Treatment Effectiveness Based on Medication Hierarchy – The Myocardial Infarction Case Study

Yingxue LI[a,1], Yiying HU[b], Jingang YANG[c], Xiang LI[a], Haifeng LIU[a], Guotong XIE[a], Meilin XU[d], Jingyi HU[d], Yuejin YANG[c,1]
[a] *IBM Research – China, Beijing, China*
[b] *Stevens Institute of Technology, New Jersey, USA*
[c] *Department of Cardiology, Fuwai Hospital, National Center for Cardiovascular Diseases, Beijing, China*
[d] *Pfizer Investment Co. Ltd., Beijing, China*

Abstract. Treatment effectiveness plays a fundamental role in patient therapies. In most observational studies, researchers often design an analysis pipeline for a specific treatment based on the study cohort. To evaluate other treatments in the data set, much repeated and multifarious work including cohort construction, statistical analysis need to be done. In addition, as treatments are often with an intrinsic hierarchical relationship, many rational comparable treatment pairs can be derived as new treatment variables besides the original single treatment one from the original cohort data set. In this paper, we propose an automatic treatment effectiveness analysis approach to solve this problem. With our approach, clinicians can assess the effect of treatments not only more conveniently but also more thoroughly and comprehensively. We applied this method to a real world case of estimating the drug effectiveness on Chinese Acute Myocardial Infarction (CAMI) data set and some meaningful results are obtained for potential improvement of patient treatments.

Keywords. Treatment effectiveness, medication hierarchy, myocardial infarction

1. Introduction

Treatment effectiveness is referred as the treatment's effect in the real world of medical practice. Treatment effectiveness helps clinicians to judge whether their therapies helps to improve the patients' health. Thus, how to evaluate the treatment effectiveness correctly and comprehensively is so much relevant to the intervention of the disease progress conditions.

In many observational studies, researchers design an analysis pipeline for some specific treatments they care about. Using drugs as an example, many studies focus on whether some classical drugs are effective or helpful to treat the patients. However, there are a mass of drugs can be used even towards curing one kind of symptom. Typical kinds of drugs only take a very small percent of the whole picture and the drugs are often with

[1] Corresponding author, E-mail: lyxlibj@cn.ibm.com, yangyjfw@126.com.

intrinsic hierarchical relationships. To dig out which kind of treatment is best for curing the patients' disease, thorough mining is very necessary, which means multiple treatments should be evaluated instead of only some typical ones. Thus, an automatic approach to explore all the rational treatments or comparable treatment pairs given a data set is very necessary. The approach proposed in our paper offers a way to achieve this function.

We demonstrate our approach on a real world case study of drug effectiveness analysis on Chinese Acute Myocardial Infarction (CAMI) data set. The CAMI data set is from the national CAMI Registry project which aims to provide a long-term platform for clinical research and achieve more knowledge of AMI patients by real world evidence [1]. The CAMI Registry started in the year of 2013 and 26,103 patients with AMI were registered until 2014. After data quality control and data selection, our paper includes 18744 patients and 104 features including clinical features, treatments and outcome finally. In the past years, the effect of some typical therapies such as statin [2-3], diuretic [4], beta blockers [5], etc. for the acute myocardial infarction (AMI) have been analyzed by many studies. However, most of the research only focus on a single treatment variable. Given the pre-defined medication hierarchy, our approach gives a more thorough and comprehensive assessment for the effect of different kinds of medication intervention for the AMI patients.

2. Methods

The workflow of our approach is illustrated in Figure 1. Two kinds of inputs are needed here. One is the data set of patients' records and the other is the information of the pre-defined treatment hierarchy. The first step of our approach will generate a rational target treatment based on the input hierarchy. Then the target cohort will be selected from the whole study population according to the target treatment. Several sub-steps are recruited in the third step which mainly complete the confounding reduction and effectiveness evaluation function. Details will be described in the following text.

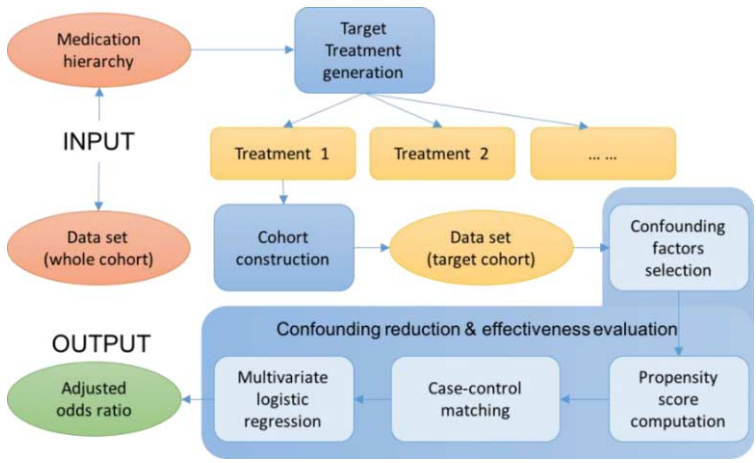


Figure 1. Workflow of our automatic approach for analyzing treatment effectiveness based on medication hierarchy.

2.1. Target Treatment Generation

The highlight part of our approach is the target treatment generation step. In this section, we will use medications in the CAMI data as an illustration to introduce the way we generate the target treatment. There are three levels in the medication information hierarchy including pharmacy class, pharmacy subclass and ingredient. Figure 2 shows different types of medication and their associated hierarchy information existing in the CAMI data set.

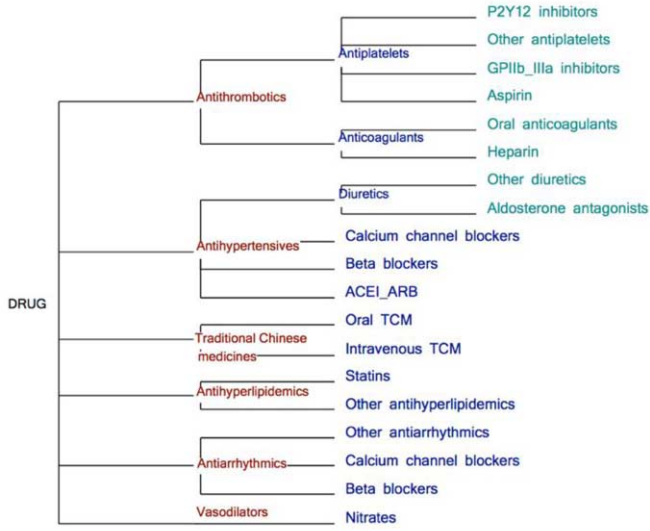


Figure 2. Different types of medication and their hierarchy information in the CAMI data set.

For the drug nodes whose parent node is the root in the hierarchy, only one kind of treatment variable named *treatment_1vs0* will be generated by our approach where the case group are patients who take the drug while the control group are patients who don not take it.

For other drug nodes in the hierarchy, other three types of treatment variables will be derived besides the *treatment_1vs0*. Using *aspirin* as an example which is shown in Figure 3, we will generate three new target treatments for this single node. They are taking the *aspirin* versus not (Type 1), taking *aspirin* versus taking another kind of drug belonging to *antiplatelet* subclass (the parent node of *aspirin* in the Figure 2) such as *P2Y12 inhibitor* (Type 2), taking the *aspirin* versus taking any of the other drugs under the *antiplatelet* subclass (Type 3).

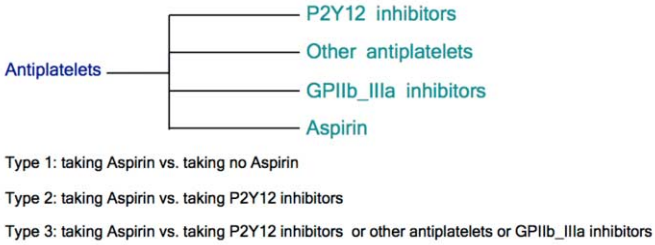


Figure 3. Three types of target treatments derived from Aspirin

2.2. Cohort Construction

Once the treatment variable is defined, the target patient cohort will be determined accordingly. It is worth mentioning that while generating new treatment variables, they can be linearly dependent on the original variables. Thus, our approach deletes those columns corresponding to that condition. For the drug included in the target treatment generation, their parent nodes and children nodes will be deleted from the data set.

2.3. Confounding Reduction & Effectiveness Evaluation

In the effectiveness calculation step, our approach uses the univariate logistic regression where outcome is the dependent variable and the treatment we derive is the independent variable as the baseline assessment.

As most of the real world cases are observational studies, confounding reduction procedure should be recruited to avoid selection bias. Our approach first using feature selection method to choose the possible confounding factors existing among the exposures. Next, we compute the propensity score [6] for each patient based on the confounding factors utilizing *MatchIt* package [7] in R. Then confounding reduction can be achieved by case control matching by the propensity score. After the matching procedure, some patients in the control group will be excluded and a new cohort will be generated. Finally, the adjusted odds ratio of the multivariate logistic regression with the propensity score and treatment as the independent variable is output as the treatment effectiveness.

3. Result

Our approach totally generates 53 rational treatment variables on the CAMI data set where twenty-five in type 1, nineteen in type 2 and nine in type 3. Table 1 lists some typical results we get. For example, our result shows that the diuretics is a significantly risk factor by baseline method but becomes significantly protective by the confounding reduction method. The confounding factors selected by our method under this scenario are sex, hypertension, the hospital patients live and so on. Thus the result makes sense as these variables also have possibilities to infect the outcome of AMI patients. The antihypertensive is a significantly protective factor by the two methods. In the pairwise drug comparisons, oral TCM seems more protective than intravenous TCM and beta blockers seems more protective than other antiarrhythmic. These results we obtained here are valuable for potential improvement of patient treatments, and could be further verified by performing specific clinical trials.

Table 1. Typical effectiveness result of the drugs on CAMI data set

Single drug effectiveness				
	Baseline		Confounding reduction	
Treatment variables	OR	p-value	OR	p-value
taking Beta blockers vs. not taking Beta blockers	0.301	<0.001	0.681	<0.001
taking statin vs. not taking statin	0.317	<0.001	0.72	<0.001
taking Diuretics vs. not taking Diuretics	1.733	<0.001	0.791	0.004
taking Antihypertensive vs. not taking Antihypertensive	0.416	<0.001	0.679	<0.001

taking Antihyperlipidemic vs. not taking Antihyperlipidemic	0.317	<0.001	0.732	0.001
pairwise drug comparisons in the same class/subclass				
Oral TCM vs. Intravenous TCM	0.404	0.003	0.534	0.055
Beta Blockers vs. Other antiarrhythmic	0.282	<0.001	0.738	0.066
drug vs. other drugs in the same class/subclass				
ACEI_ARB vs. DrugsInSameGroup	0.324	<0.001	0.747	0.080

4. Discussion

The pipeline we designed here offers a way to automatically explore the treatment effectiveness from the given data set in batch. Although we use drugs in the CAMI data set as an illustration, our approach can be easily generalized to other kinds of treatments. For example, the treatment trajectory can be viewed as a treatment type. The convenience of our approach can be more obvious on it because evaluation would be very computational complex in the manual way. In addition, patients clustering procedure can also be recruited to our approach to gain a more accurate assessment. Via clustering procedure, we categorize the whole patient cohort into several subgroups according to their nature characteristics. Then treatment effectiveness can be evaluated further on each sub-groups. Multiple testing should also be done to the p-values of the odds ratio for further improvements of our pipeline. [8] More precise result could be got from this.

References

[1] Xu H, Li W, Yang J, et al. The China Acute Myocardial Infarction (CAMI) Registry: A national long-term registry-research-education integrated platform for exploring acute myocardial infarction in China[J], American heart journal. 2016; 175: 193-201. e3.
[2] Spencer FA, Fonarow GC, Frederick PD, Wright RS, Every N, Goldberg RJ, Gore JM, Dong W, Becker RC, French W, early withdrawal of statin therapy in patients with non–st-segment elevation myocardial infarction national registry of myocardial infarction, Arch Intern Med. 2004; 164(19):2162-2168.
[3] Tonkin A, Alyward P, Colquhoun D, et al. Prevention of cardiovascular events and death with pravastatin in patients with coronary heart disease and a broad range of initial cholesterol levels[J], New England Journal of Medicine. 1998; 339(19): 1349-1357.
[4] Abraham AS, Rosenman D, Meshulam Z, Balkin J, Zion M, Eylath U. Intracellular cations and diuretic therapy following acute myocardial infarction, arch intern med. 1986; 146(7):1301-1303.
[5] Hjalmarson Å, Herlitz J, Malek I, et al. Effect on mortality of metoprolol in acute myocardial infarction: a double-blind randomised trial[J], The Lancet; 1981; 318(8251): 823-827.
[6] Austin P C. An introduction to propensity score methods for reducing the effects of confounding in observational studies[J], Multivariate behavioral research; 2011; 46(3): 399-424.
[7] Ho D E. MatchIt: nonparametric preprocessing for parametric causal inference[D]. Departments of Mental Health and Biostatistics, Johns Hopkins Bloomberg School of Public Health, 1737.
[8] Abdi, Hervé. Bonferroni and Šidák corrections for multiple comparisons, Encyclopedia of measurement and statistics. 2007;3:103-7.

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-111

Evaluation of Machine Learning Methods to Predict Coronary Artery Disease Using Metabolomic Data

Henrietta FORSSEN[a*1], Riyaz PATEL[b,c*], Natalie FITZPATRICK[b,c], Aroon HINGORANI[d], Adam TIMMIS[e], Harry HEMINGWAY[b,c], Spiros DENAXAS[b,c]
[a]*Department of Computer Science, UCL*
[b]*Institute of Health Informatics, UCL*
[c]*Farr Institute of Health Informatics Research, UCL*
[d]*Institute of Cardiovascular Sciences, UCL*
[e]*NIHR Cardiovascular BRU, Barts*

Abstract. Metabolomic data can potentially enable accurate, non-invasive and low-cost prediction of coronary artery disease. Regression-based analytical approaches however might fail to fully account for interactions between metabolites, rely on a priori selected input features and thus might suffer from poorer accuracy. Supervised machine learning methods can potentially be used in order to fully exploit the dimensionality and richness of the data. In this paper, we systematically implement and evaluate a set of supervised learning methods (L1 regression, random forest classifier) and compare them to traditional regression-based approaches for disease prediction using metabolomic data.

Keywords. coronary artery disease, random forest, machine learning, EHR

1. Introduction

Coronary artery disease (CAD) is one of the leading causes of morbidity and mortality worldwide [1]. Definitive diagnosis is by coronary angiography, an invasive procedure that can lead to severe complications [2] or by additional often costly imaging techniques. Non-invasive blood testing, using circulating metabolites [3] [4], could potentially minimize unnecessary tests and predict CAD with higher accuracy. Previous research however has been mainly restricted to classical regression-based methods [5] [6] and potentially fails to fully exploit the dimensionality and richness of the data by accounting for interactions between metabolites. Supervised machine learning (ML) methods might be better-placed to address these challenges but have yet to be systematically evaluated in this context. Our aims were to a) investigate and evaluate supervised ML methods for CAD prediction using metabolomics data and b) compare their accuracy with traditional regression-based approaches.

[1] Henrietta Forssen, Department of Computer Science, UCL, London, United Kingdom; E-mail: henrietta.forssen.15@ucl.ac.uk.

2. Background

Metabolites are small molecules produced during metabolism or generated by microbes within the body [7]. Metabolites are the end-products of gene expression, a process closely related to protein/enzymatic reactions and therefore potentially offer a direct molecular reflection of the cellular milieu that leads to pathophysiological changes. Circulating metabolites may help predict the presence of CAD by firstly identifying metabolic disturbances, relevant for atherosclerosis (e.g. diabetes and insulin resistance [8] [9]). Additionally, since atherosclerosis occurs at the blood-vessel wall interface, blood metabolite measurements could plausibly directly reflect this chronic process and help predict CAD existence and stability. However, for selected metabolites studied to date, the incremental predictive utility over routine clinical assessments has been modest and restricted to a few candidates measured using non-scalable methods. Recent high-throughput, low-cost and high-dimensional methods [4] (e.g. nuclear magnetic resonance spectroscopy), have re-invigorated hope for using metabolic signatures for CAD prediction but analyses of these complex data present new challenges before realized.

ML techniques are data-driven approaches designed to discover statistical patterns in large high-dimensional multivariate data and have been previously used for creating accurate risk prediction models [10]. Supervised ML methods are a set of techniques which aim to infer a function from a labelled training dataset which can predict the class of future input vectors. We evaluated penalized logistic regression and random forest to assess the predictive performance of metabolites on CAD in a contemporary cohort of patients referred to hospital for chest pain investigation or planned coronary angiography.

3. Methods

3.1. Case and Exposure Definitions

We used data from the Clinical Cohorts in Coronary disease Collaboration (4C) study (n=3409) which recruited patients with acute or stable chest pain from four UK NHS hospitals [11]. Patients consented to having their EHR extracted and provided blood samples. We defined presence of CAD as a >50% stenosis [12], occurring in ≥1 coronary arteries using data from: a) coronary angiography reports and b) EHR evidence of previous coronary revascularization procedures (Percutaneous Coronary Intervention, Coronary Artery Bypass Graft) recorded in EHR. Participants in whom CAD could not be ascertained were excluded. For each participant, 256 metabolites were quantified using an NMR technique. Full details have been published elsewhere [4] [11]. Missing metabolite values were imputed and zero mean unit standardized by multiple imputation [13] (predictive mean matching [14]) and standardized to zero mean unit variance by first subtracting the means and dividing by the standard deviations. Data were randomly split into training and test subsets using a 3:1 ratio.

3.2. Statistical Methods

We performed logistic regression on each of log+1-transformed metabolite values adjusting for known risk factors. We derived principle component factors for the

standardized metabolite values and selected the first six for analyses as they accounted for >95% of the data variability. We then performed logistic regression on each of the Principal Component Analysis (PCA)-derived metabolite factors, and multiple logistic regression including all six. Adjusted (age, sex, use of statins, hypertension) and unadjusted models were Bonferroni corrected ($p<0.05$). We performed penalized logistic regression using the Lasso penalty which was defined as the lowest error obtained from a 50-fold cross-validation. We trained a random forest classifier using Gini impurity and 5,000 trees per ensemble. Initial cross-validation was conducted on the training set for both the proportion of variables used per tree as well as the maximum tree depth. A second cross-validation was conducted on the number of variables alone, whilst allowing trees to grow to their maximum depth. This removed the uncertainty of tuning a second parameter, and the possible increase in variance due to increased depth was considered well counterbalanced by using a very large number of ensemble trees. Final predictions were the average individual pooled predictions [13] [14] across imputed datasets and evaluated by calculating the percentage of correct predictions, ROC curves and AUC.

4. Results

We identified 1474 patients with metabolomics in whom CAD was ascertained (Table 1).

Table 1. Summary of study population

Clinical Characteristics		Clinical Characteristics	
Men (%)	1106 (78%)	Statin use (%)	447 (30%)
Age (Years)	62.4± 11.6	Diabetes (%)	523 (35%)
BMI (kg/m2)	29.4± 5.3	Current smoker (%)	278 (28%)
Diagnosed hypertension (%)	1146 (77%)	CAD present (%)	1037 (70%)

4.1. Comparison of Model Predictiveness

In the unadjusted models, the random forest classifier had the highest AUC and accuracy values and highest ROC curve (Figure 1) and both ML models outperformed PCA regression (Table 2). AUC, raw accuracy and PPV were mostly similar across models. All models had significantly higher sensitivities than specificities, but PCA regression had the most extreme values as it predicted the vast majority of positive CAD cases correctly but nearly none of the negative CAD cases. The large disparity in sensitivity and specificity for the two other two models shows that they failed to accurately distinguish between disease states. When adjusting for confounders, PCA regression had the best accuracy but higher AUC and accuracy values compared to unadjusted models were observed in all models.

Table 2. Adjusted/unadjusted prediction results; highest AUC values highlighted.

Model	Accuracy	AUC	Sensitivity	Specificity	PPV	NPV
PCA regression	0.686	0.625	0.984	0.026	0.691	0.429
PCA regression adjusted	0.759	0.767	0.957	0.322	0.757	0.771
L1 regression	0.688	0.663	0.882	0.261	0.725	0.550
L1 regression adjusted	0.767	0.765	0.949	0.339	0.760	0.750
Random forest	0.713	0.675	0.941	0.209	0.724	0.615
Random forest adjusted	0.732	0.711	0.937	0.278	0.741	0.667

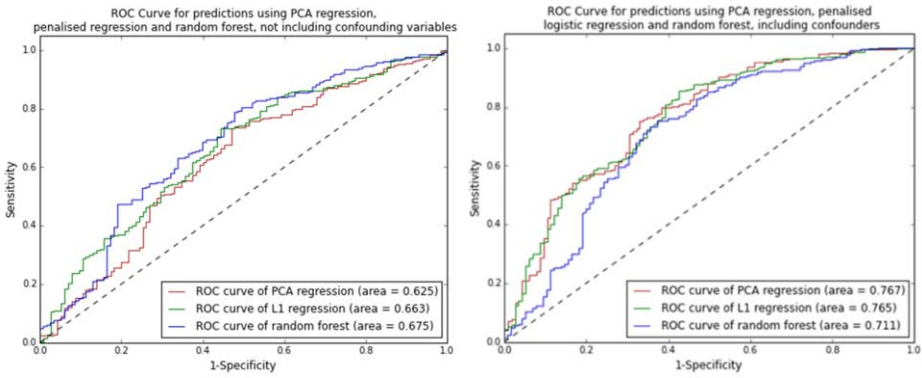


Figure 1. Unadjusted (left) and adjusted (right) ROC curve

4.2. Model Compositions and Predictive Metabolites

Logistic regression: In individual metabolites, strong CAD associations after Bonferroni correction were primarily related to lipids, (e.g. HDL and VLDL), apolipoprotein A-I, the ratio of triglycerides to phosphoglycerides and the ratio of omega-6, monosaturated, polysaturated values to total fatty acids. With logistic regression on individual PCA-derived factors, the first PCA factor (VLDL, ratio of apolipoprotein B to apolipoprotein A-I, ratio of apolipoprotein B to apolipoprotein A-I, 0.404 variance) remained statistically significant. When using all factors and adjusting for confounders, the first and second factors (IDL and LDL, 0.165 variance) were statistically significant.

Penalized logistic regression: In the unadjusted models, the ratio of apolipoprotein B to apolipoprotein A-I was found to have the largest, statistically significant negative association with presence of CAD. This was followed by cholesterol esters in small LDL which had a large positive association with presence of CAD. Saturated fatty acids also had a large negative association, whilst phospholipids in chylomicrons and extremely large VLDL had a large positive association. Overall ~70 predictive metabolites where included in each model with 117 metabolites included in at least one of the models used to average the prediction. This suggests that whilst excluding confounders, it is difficult to select a small profile of metabolites to accurately predict the presence of absence of CAD using penalized regression. When adjusting for confounders, substantially fewer metabolites were selected; the ratio of monounsaturated fatty acids to total fatty acids and triglycerides to total lipids ratio in IDL had the largest statistically significant positive association while glutamine and acetoacetate had negative associations.

Random forest: In the unadjusted classifier, creatinine was the most strongly significant metabolite followed by triglycerides to total lipids ratio in IDL, phenylalanine, albumin and lactate. Similar predictors were observed in the adjusted models with age being the most significant component followed by creatinine, triglycerides to total lipids ratio in IDL, phenylalanine, albumin and lactate. Similar metabolite profiles for adjusted/unadjusted models suggest that random forest does not incorporate the additional information of confounding variables as well as the other models.

5. Concluding Discussion

While ML approaches predicted presence/absence of CAD in the unadjusted models (using metabolite data only) with high accuracy/sensitivity, when adjusting for confounders they were outperformed by PCA regression in terms of ROC AUC and accuracy suggesting that a small number of metabolites can potentially be included in prediction models. Multiple individual metabolites that were found statistically significant are in agreement with previous literature and our pathological understanding of CAD and its development. Among these, the atherogenic lipid particles such as LDL are known to be causally related to atherosclerosis, while others such as creatinine reflect renal function and are also established markers of CAD risk. Several other metabolites have no previous robust association with CAD including phenylalanine and lactate and represent potentially novel avenues for investigation. However in seeking a metabolic signature to predict CAD, ML models suffered from low specificity.

This exploratory analysis has identified and exemplified the value of ML models for CAD prediction using high-dimensional data, and shown that accuracy of traditional regression-based approaches can be surpassed. Nonetheless further research is required before these methods can be translated into clinical solutions.

References

[1] Scarborough PB, Wickramasinghe K, et al., *Coronary heart disease statistics*, British Heart Foundation, 2010.
[2] NICE, *Chest pain of recent onset, Clinical Guidance CG95*, 2010.
[3] Mayr M, Metabolomics:ready for the prime time?, *Circulation: Cardiovascular Genetics* **1** (2008), 58-65.
[4] Soininen P, Kangas A, Würtz P, et al., Quantitative Serum Nuclear Magnetic Resonance Metabolomics in Cardiovascular Epidemiology and Genetics, *Circulation: Cardiovascular Genetics* **8** (2015), 192-206.
[5] Würtz P, Havulinna A, Soininen P, et al., Metabolite Profiling and Cardiovascular Event Risk: A Prospective Study of Three Population-Based Cohorts, *Circulation* **131** (2015), 774-785.
[6] Kirschenlohr H, Griffin J, Clarke S, et al., Proton NMR analysis of plasma is a weak predictor of coronary artery disease, *Nature Medicine* **12** (2006), 705 - 710.
[7] Dr. Villas-Bôas S, Dr. Roessner U, Dr. Hansen M, et al., *Metabolome Analysis: An Introduction*, Wiley, 2006.
[8] Würtz P, Mäkinen VP, Soininen P, et al. Metabolic Signatures of Insulin Resistance in 7,098 Young Adults, Diabetes **61** (2012), 1372-1380.
[9] Wang T, Larson M, et al. Metabolite profiles and the risk of developing diabetes, *Nat Med* **17** (2011), 448–453.
[10] Goldstein B, Navar AM, Carter R, Moving beyond regression techniques in cardiovascular risk prediction, *EHJ* **(**2016), 10.1093/eurheartj/ehw302 .
[11] Hemingway H, Feder GS, Fitzpatrick NK, et al, Using nationwide “big data” from linked EHR to help improve outcomes in cardiovascular diseases CALIBER., NIHR PGfAR (2016) (in-press).
[12] Harris PJ, Behar VS, Conley MJ, et al. The Prognostic Significance of 50% Coronary Stenosis in Medically Treated Patients with Coronary Artery Disease, *Circulation* **62** (1980), 240-248.
[13] Rubin DB, *Multiple Imputation for Nonresponse in Surveys,* Wiley, New York, 1987.
[14] van Buuren S, Groothuis-Oudshoorn K, mice: Multivariate Imputation by Chained Equations in R, *Journal of Statistical Software* **45** (2011)

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-116

Dermatology Disease Prediction Based on Two Step Cascade Genetic Algorithm Optimization of ANFIS Parameters

Aja AVDAGIC [a] and Lejla BEGIC FAZLIC[b,1]
[a] *Faculty of Medicine-Ludwig Maximilian University of Munich*
[b] *University of Sarajevo – Faculty of Electrical Engineering*

Abstract. The aim of this study is to present novel algorithms for prediction of dermatological disease using only dermatological clinical features and diagnoses collected in real conditions. A combination of the Adaptive Neuro-Fuzzy Inference Systems (ANFIS) and Genetic algorithm (GA) for ANFIS subtractive clustering parameter optimization has been suggested for the first level of fuzzy model optimization. After that, a genetic optimized ANFIS fuzzy structure is used as input in GA for the second level of fuzzy model optimization. We used double 2-fold Cross validation for generating different validation sets for model improvements. Our approach is performed in the MATLAB environment. We compared results with the other studies. The results confirm that the proposed model achieves accuracy rates which are higher than the one with the previous model.

Keywords. Adaptive Neuro-Fuzzy Inference System, Genetic Algorithm, Cross Validation, Prediction, Dermatological Diseases

1. Introduction

Dermatology is a study of skin disease that is very complex and difficult to diagnose, and ultimately may be a leading cause of skin cancer. Differential diagnostic procedure is a systematic diagnostic method used to identify the presence of a disease entity where multiple alternatives are possible. This method is essentially a process of elimination or at least a process of obtaining information that shrinks the "probabilities" of candidate conditions to negligible levels, by using evidence such as symptoms, patient history, and medical knowledge. The five different categories: pityriasis, seboreic dermatitis, lichen planus, pityriasis rosea and cronic dermatitis have been observed in this study. They all share the clinical features of erythema and scaling with very few differences [1] so usually it is difficult to identify the particular diseases present in a patient. Numerous authors contributed various data mining algorithms for the diagnosis of dermatology diseases. H.Altay Guvenir et.al [2] has proposed a new classification algorithm VFI5 that is voting feature interval and has achieved 96.25% accuracy.

Our previous research [1] presented a new approach based on ANFIS model for the detection and recognition in different types of dermatological diseases where five ANFIS classifiers were used to detect different types of dermatological diseases. Each of the ANFIS classifiers was trained so that they are likely to be more accurate for one class of

[1] Lejla Begic Fazlic, Aleja lipa 53, 71 000 Sarajevo, Bosnia and Herzegovina; E-mail: lejla.begic@fds.ba.

disease than for the others. Authors in paper [3] have developed a hybrid model based on SVM and artificial neural network. G. Castellanoe et al. present in their recent work [4] the application of a particular neuro-fuzzy system, named KERNEL, to the problem of differential diagnosis of erythemato-squamous diseases. In study [5] is shown a hybrid model based on multilayer perceptron, decision tree and LDA. Our recent study [6] presents a GA-ANFIS expert system prototype for prediction of dermatological diseases and has achived 97,8 % accurancy. Proposed algorithms in this study, aim to predict the dermatology diseases with highest accuracy using only clinical features without biopsy results. The results are compared with other studies and they show the effectiveness of the proposed approach.

The paper is set in three parts. The introduction part describes problems about differential diagnosis of dermatological diseases. We have listed some research studies on artificial intelligence approaches. In part two we have explained the methodological framework used in our research along with the description of proposed algorithm and double 2-fold validation process. We have described the proposed system design architecture using the MATLAB user interface. In this part are also shown the numerical training results and test error compared with the other results. Finally, the discussion and conclusion is summarized in part three.

2. Methodological Framework and Results

The diseases observed in this group are: pityriasis, seboreic dermatitis, lichen planus, pityriasis rosea and cronic dermatitis. All observed diseases share the clinical features of erythema scaling with very few differences [2]. Data base investigated in this study consisted of 345 analysed data cases [7]. 285 analysed patents are used for training and checking, and one independent data set consisting of 60 individuals is used for model validation. We used cross validation on 2-folding levels to obtain 27 different data sets (for training, test and validation) as is presented in Figure 1. The nine input attributes are: erythema, scaling, definite borders, itching, Koebner phenomenon, polygonal papules, follicular papules, oral mucosal involvement and knee and elbow involvement. We used three different statuses for features (3-obviously present, 0-not present, 1-2 intermediately presented values). In earlier approach we used triple ANFIS method [8] based on Sugeno models using fuzzy logic and fuzzy sets. Each of them is based on skin features due to chromosome structure limitation length and processor memory limitation. The Sugeno model we used has three inputs (skin features), one output (type of diseases), and 27 IF-THEN knowledge base rules.

The novel proposed algorithm uses all 9 features at once. We used 27 different data that we created using double 2-cross validation (one set for training, one set for test, and one set for validation) as is presented in Figure 1 as inputs in the ANFIS model. The GA based on evolutionary paradigm (crossover, mutation, population, generation and fitness function) is used in this study. The main function of GA is to share four ANFIS subtractive parameters structure and then optimizes fuzzy sets in knowledge rule base with the aim of arriving at the best prediction of dermatological diseases.

In the first step, we use a novel algorithm to make genetic optimization of ANFIS subtractive clustering parameters (range of influence, squash ratio, accept ratio, reject ratio) resulting in the best subclustering points. The size of independent variables for the fitness function is 4 and chromosome structure is generated by ANFIS subtractive clustering parameter. After applying the best subtractive clustering parameter, each of

the obtained clusters will constitute a prototype for a particular behavior of the system under analysis. So, each cluster can be used to define a fuzzy rule capable of describing the behavior of the system in some region of the input–output space. This genetic optimized fuzzy structure (we denote it as GA ANFIS SC*) is used as an input in the second phase, where we used GA to optimize GA ANFIS SC* fuzzy structure on membership functional level which resulted in the best prediction of dermatological diseases. The size of chromosome structure-fuzzy structure on membership functional level was generated by ANFIS computational complexity [9]. Proposed novel GA GA ANFIS algorithm is described in Figure 1.

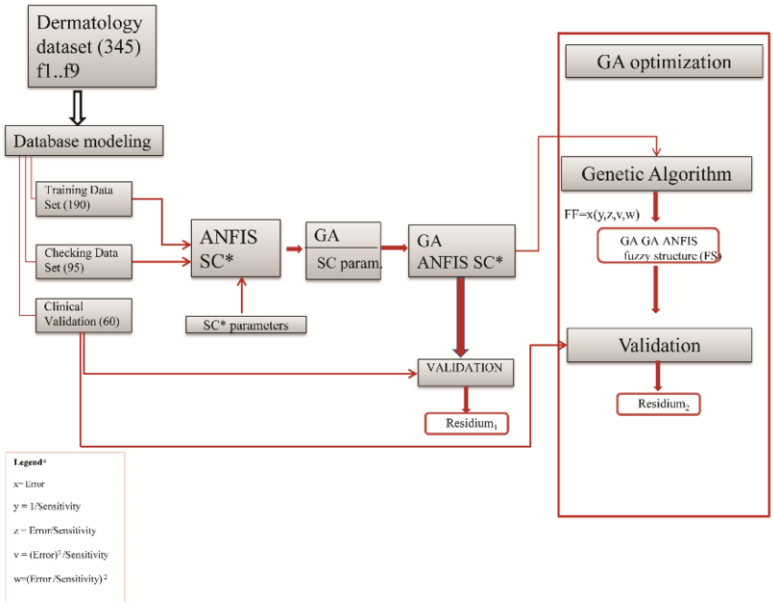


Figure 1. Proposed model: GA GA ANFIS algorithm

We needed an additional validation of the system that goes to the application, which we achieved with different validation datasets (60 separated patients) as is presented in Figure 2.

The Matlab GUI (Figure 3 (a)) consists of five parts. The first part contains pre-processing dermatology data. The second part is used for genetic optimization of 4 ANFIS subtractive clustering parameters. This part is also aimed at creating of GA ANFIS SC* fuzzy structure (with best (GA) SC parameters). Third part is used for GA parameterization (population size, number of generations, GA operators). The fourth part is aimed at genetic optimization of GA ANFIS SC* structure. The last, fifth part integrates all the previous data and functions into operations with a disease class and an error prediction. We made validation of prototype through 27 experiments. The ANFIS training error, statistical measure of performance (Sensitivity, Error/Sensitivity, Error^2 /Sensitivity and (Error /Sensitivity)^2) were used as the fitness criterion in the evaluation function. Finally, Figure 3 (b) shows the comparison of the predicted value versus target value of the validation data for the best model.

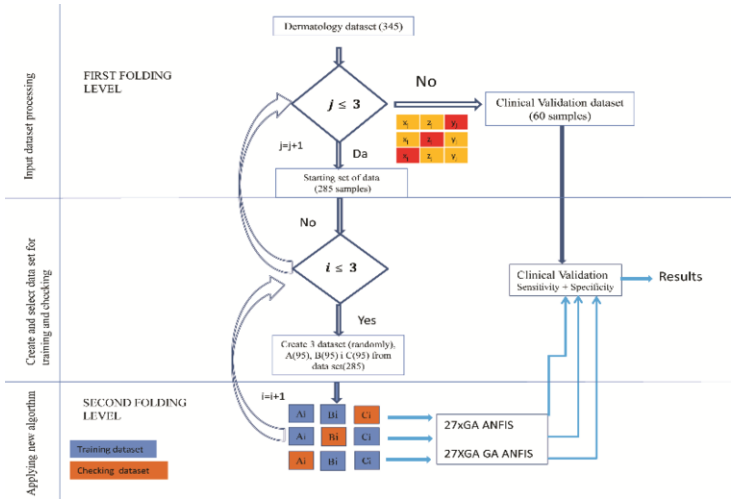
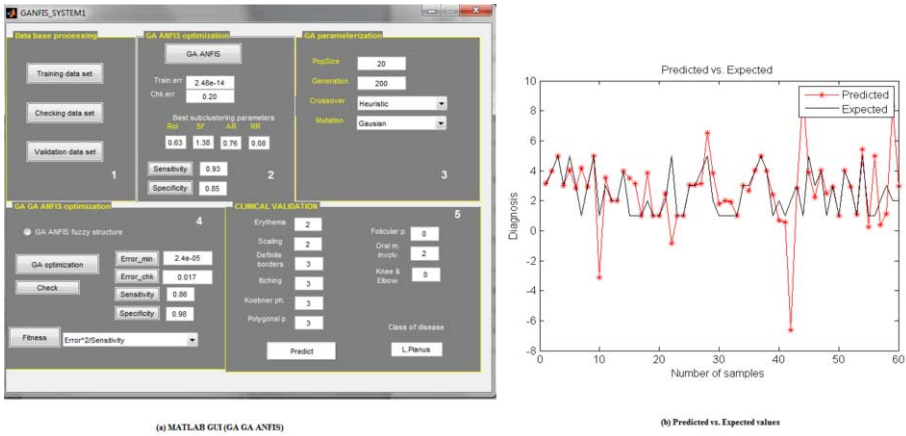


Figure 2. Double 2-fold cross validation



(a) MATLAB GUI (GA GA ANFIS)

(b) Predicted vs. Expected values

Figure 3 (a) GA GA ANFIS GUI and (b) Predicted vs. Expected values

Table 1 shows the comparison results of the validation success of proposed algorithms (GA ANFIS and GA GA ANFIS) versus the other ones previously used as follows:

Table 1. Compared results – proposed method vs. other studies –Dermatology data set

Authors	Methods	GA ANFIS	GA GA ANFIS
Proposed method in this study	**(double 2-Fold Cross validation) 27 data sets, GA-GA-ANFIS**	Train Err: 2.4e-14 Check Err: 0.207 Sensitivity:0.97 Specificity:0.93 TA:96.6	Train Err: 8.0-05 Check Err: 0.003 Sensitivity:1 Specificity:0.93 TA:98.66
Begic L.et al [6]	**GA-FUZZY**	TA:97.89	
Amarathunga. et al. [10]	**GA,SVM**	TA:85%	
Barati E. et al. [3]	**Machine learning**	TA: 92,5%	
Begic Fazlic L. et al. [1]	**ANFIS**	TA: 95,56	
Pappa G.L. et.al [11]	**GA-MOGA**	TA: 76,5-86,5	

3. Conclusion

The comparison results show that the proposed algorithm based on two step cascade GA optimization has a better performances than the ones in [1][3][6][10][11]. Proposed model runs in two steps.

In the first step algorithm combines double 2-fold cross validation to generate different data sets for training, test and validation (27 data sets).

In the next step, ANFIS and GA capabilities are used to generate best subtractive clustering parameters producing GA ANFIS SC* structure which is used in GA optimization for different values of fitness function. Results show that the proposed model can be used in detecting classes of dermatological diseases by taking into consideration only clinical features.

References

[1] Z. Avdagic, L.B. Fazlic, H. Fatkic, Automatic detection of dermatological diseases using Adaptive Neuro Fuzzy Inference Systems, Acta Informatica Medica, Journal of Society for Medical Informatics of B&H, Vol. 15 Issue 2; (2007).

[2] E. Derya Ubeyli, I. Guler, Automatic detection of erythemato-squamous diseases using adaptive neuro fuzzy inference systems, Elsevier -Computers in Biology and Medicine 35 (2005), 421–433

[3] E. Barati, M. Saraee, A. Mohammadi, N. Adibi, M.R. Ahamadzadeh, A survey on implementation of machine learning techniques for dermatology diseases classification, Journal of Selected Areas in Health Informatics (JSHI), (2011).

[4] G. Castellano, A.M. Fanelli, C. Leone, Diagnosis of dermatological diseases by neuro fuzzy system, Conference: Proceedings of the 3rd Conference of the European Society for Fuzzy Logic and Technology, Zittau, Germany, September 10-12 (2003).

[5] A.M. Elsayad, Diagnosis of erythemato-squamous diseases using ensemble of data mining methods, ICGST-BIME Journal, 10(1) (2010), 13-23.

[6] L.B. Fazlic, K. Avdagic, S. Omanovic, GA-ANFIS Expert System Prototype for Prediction of Dermatological Diseases, Stud Health Technol Inform. (2015); 210, 22-6.

[7] https://archive.ics.uci.edu/ml/datasets/Dermatology

[8] J.-S.R. Jang, ANFIS: Adaptive-network-based fuzzy inference system, IEEE Trans. Syst. Man Cybern. 23 (3), (1993), 665–685.

[9] P. Bonissone, Adaptive Neural Fuzzy Inference System, Analysis and Application, http://homepages.rpi.edu/~bonisp/fuzzy-course/Papers-pdf/anfis.rpi04.pdf.

[10] A. Amarathunga, E. Ellawala, G.N. Abeysekara, C.R.J. Amalraj, Expert System For Diagnosis Of Skin Diseases, International Journal of Scientific & Technology research, Volume 4, Issue 01 (2015).

[11] G.L. Pappa, A. Alex, F. Celso A.A., A.A. Kaestner, A Multiobjective Genetic Algorithm for Attribute Selection, DBLP: 16.SBIA 2002.

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-121

Querying EHRs with a Semantic and Entity-Oriented Query Language

Romain LELONG[a], Lina SOUALMIA[a,b], Badisse DAHAMNA[a], Nicolas GRIFFON[a,b] and Stéfan J. DARMONI[a,b,1]
[a] *Department of Biomedical Informatics, Rouen University Hospital, France*
[b] *French National Institute for Health, INSERM, LIMICS UMR-U1142, Paris, France*

Abstract. While the digitization of medical documents has greatly expanded during the past decade, health information retrieval has become a great challenge to address many issues in medical research. Information retrieval in electronic health records (EHR) should also reduce the difficult tasks of manual information retrieval from records in paper format or computer. The aim of this article was to present the features of a semantic search engine implemented in EHRs. A flexible, scalable and entity-oriented query language tool is proposed. The program is designed to retrieve and visualize data which can support any Conceptual Data Model. The search engine deals with structured and unstructured data, for a sole patient from a caregiver perspective, and for a number of patients (e.g. epidemiology). Several types of queries on a test database containing 2,000 anonymized patients EHRs (i.e. approximately 200,000 records) were tested. These queries were able to accurately treat symbolic, textual, numerical and chronological data.

Keywords. Electronic Health Records, Information Storage and Retrieval, Search Engine, Controlled vocabulary

1. Introduction

Electronic Health Records (EHR) play a central role since they include a long-term record of care and a record of events from different types of care, including instructions, prospective information such as plans, orders and evaluations. In this context, the goal of an Information Retrieval (IR) System on EHR is to provide physicians with the correct information at the right place for the right person. Several tools and frameworks for searching in EHRs for one patient have been proposed. These tools are adapted according to each data format: structured, not structured or mixed. The main system is Informatics for Integrating Biology and the Bedside (I2B2), an open source platform developed in the USA and dedicated to translational research. The I2B2 center focuses on developing a scalable informatics framework to bridge clinical research data with basic sciences research data. The framework uses coded data, biological data and other genomic data. The scope of the search concerns clinical search and statistical data analysis. Data semantics is particularly important as it derives from the concrete healthcare providing process in hospitals. EHR data is mainly composed of several key

[1] Pr. Stéfan J. DARMONI, MD, PhD, Department of Biomedical Informatics, 1 rue de Germont 76031 Rouen Cedex, France, Cours Leschevin, Porte 21, 3ème étage, Email : Stefan.Darmoni@chu-rouen.fr

entities semantically related to one another: (a) patient, (b) hospital, (c) stay and then (d) the "classical" and more basic levels (procedures, diagnosis related group (DRG) coding, lab tests, reports, metadata from reports etc.). As a consequence, IR from EHR is more difficult and different when compared to the "classical" IR. In this context, the aim of this study was twofold. First, describe a conceptual data model (CDM) which represents the conceptual and intuitive representation that non-IT medical provider users can have of EHR data. Secondly, describe a query language (QL) used to query those data and providing users the possibility to build queries accessing the entire set of EHR entities by taking advantage of the semantic network of entities. This study has been carried out within the context of the Retrieval and Visualization In Electronic Health records (RAVEL) project.

2. Materials

EHR Data Sources: A corpus of 2,000 anonymized patients and 200,000 reports from Rouen University Hospital (RUH) was used in this study, approved by the French National Commission on Computers and Liberty. Almost any clinical information available in the EHR is integrated in the RAVEL model, e.g. DRG codes (ICD10), patient data (age, gender), lab tests and all medical reports.

EHR Conceptual schema and data model: The underlying database of the system is based on a generic Entity-Attribute-Value (EAV) physical data model [1]. This data model is able to integrate all types of data in only a few tables without structural changes to the data model (e.g. columns or tables addings). This helps to optimize IR, maintain the database and manage heterogeneous data types. A dedicated CDM was designed to abstract the EHR data contained in the physical database data model. The query language syntax is patterned on that CDM instead of the physical database schema which provides the Search Engine (SE) with semantic features and capabilities.

3. Materials

3.1. Query Language Description

The specific QL syntax is based on the CDM. Hence, building a query only requires real-life knowledge of existing entities in the database, their properties and their relationships with each other. This QL has three main characteristics:

- *Semantic IR capabilities:* The QL is built with an entity-oriented vision. It enables semantic information retrieval since it provides the ability to display and query EHRs semantically related entities on any level (patient, stay, procedure, biology etc.). It can also deal with multiple terminologies and hierarchical relationships.
- *Scalability & flexibility:* The QL automatically handles modifications on the CDM (i.e. new conceptual entities, attributes and relationships between entities) without any SE modification. This enabled an easy and rapid extension to omics data [2].
- *Comprehensive querying:* The full scope of entities can be queried using constraints built upon several types of data: Textual and symbolic data (e.g.

patient(gender="M")), Numerical data (e.g. *medicalTest(6<numericResult<= 6.25))* and Chronological (eg. *stay(entryDate>2010-03-10))*. All comparators and operators available are specified in Table 1.

Table 1. Types of data handler by the search engine

Data type	Available operators	Available comparators
Character string data	None	= (equal), != (not equal), * (wildcard)
Numerical data	+ (add), - (substract), * (multiply), / (divide)	=, !=, < (lower), <= (lower or equal), > (greater), >= (greater or equal)
Chronological data	+,-	=,!=,<,<=,>,>=

3.2. Query Language Description

Basic querying: The query language is basically composed of nested syntactical units with the following syntax *ENTITY(CONSTRAINTS_CLAUSE)*. *ENTITY* can correspond to any kind of entity of the CDM (e.g. *patient*, *stay*, *medicalUnit* etc.) and specify the type of object that the SE should return (or target when nested). For instance, the queries *patient()* and *medicalUnit()* would respectively return all the patients and all the medical units of the database. The *CONSTRAINTS_CLAUSE* is a boolean expresion enabling to apply constraints to the targeted *ENTITY*. For instance, the query *patient(birthDate=1937-01-01 AND gender="M")* uses the two attributes birthDate and gender of the patient entity to return all male patients born on *1937-01-01*. *stay(leavingDate–entryDate>=10)* will return stays with a duration of 10 days or more.

Semantic querying: The strength of the query language originates from its ability to deal with nested syntactical units. For instance the query *stay(patient(id= "DM_PAT_42"))* targets stays link to at least one relationship to the patient 42. More complex queries can be performed by using the relationships between these entities (Table 2). This nesting functionality allows the exploitation of the relationships between entities and thereby enables to build queries based on the full semantic network. The QL has other querying capabilities: full text search, minimum and maximum on numerical data, hierarchical expansion, chronological and temporal queries.

4. Result

Several use cases were successfully answered in the RAVEL project:

- *Use case 1:* Visualize over time the neutrophil rate of a patient with rheumatoid arthritis,
- *Use case 2:* Produce all the medical reports containing the concept of metastasis,
- *Use case 3:* Retrieve all stays where *"REMICADE"* (infliximab) was used.

The use cases resolution required to use: Automatic Indexing in medical records, full text search, and multiple terminological resources. Some of the queries used to answer these three use cases are shown in Table 2.

4.1. Comparison to I2B2 workbench

The I2B2 workbench and the QL described in this study are both tools designed for searching in EHRs. However, the two tools have differences which are summarized in Table 3. The I2B2 workbench provides numerous default features which cover a lot of use cases. It notably enables to detect the number of occurrences of an event contrary to the QL described in this study. The database on which the QL operates integrates currently 69 English and French terminologies which represent 2,340,655 concepts partially translated into French. I2B2 workbench includes 14 terminologies (cf. Table 3) English for the major part. Other terminologies can be added. In contrast to I2B2 workbench, reports are automatically indexed and can be queried using the terminology terms with the QL. As regards cohort patient selection, I2B2 and the QL share most of their functionalities such as: numerical, chronological and textual constraints, full-text search on reports, search using concept subsumption, use of clinical data as constraints (stay, medical unit, patient, etc.) and omic variant data management.

Table 2. Query examples

Semantic query examples	*stay(patient(id_"DM PAT 1736") AND medicalUnit(label="Cardiology"))* All the patient 1736 stays which occur in the Cardiology medical unit.
	stay(icd10SC(label="Burns involving less than 10% of body surface")) stays with a diagnosis of *Burns involving less than 10% of body surface* (T31.0 sub category of ICD10).
	medicalTest(medicalTest(label="Sodium") AND numericResult<lowerBound AND patient(id="DM PAT 125")) For a given patient (number 125), display all hyponatremia test results.
	patient(stay(icd10SC(id = "CIM SC T31.0") AND medicalTest(exe(label="Sodium") AND numericResult>upperBound))) patients coded with the T31.0 sub category of ICD10 DRG code showing hypernatremia in that stay.
RAVEL queries	*stay(patient(id="DM PAT 21") AND procedure(label="BLOOD SAMPLE"))* Patient 21 stays in which a blood sample procedure was performed.
	medicalUnit(stay(patient(id="DM PAT 21") AND procedure(label="BLOOD SAMPLE"))) Medical units of the patient 21 stays in which a blood sample was taken.
	*biologicalTest(patient(id="DM PAT 1078") AND exe(label="Platelets") AND 10*numericResult<lowerbound)* Patient 1078 platelet tests with a result more than 10 times lower than normal level.
	procedure(ccamMP(id="CCA AM EQQM006") AND procedureDate="MAX") The last procedure coded with EQQM006.

Table 3. QL vs I2B2 Functionalities

	QL	I2B2
Querying scope	1 or n entity	n patients
Querying	Textual query	Graphical query
Detection of number of event occurrences	NO	YES
Lab test unit choice	NO	YES
Defaultly supported terminologies	69	14
Record Automatic Indexing	YES	NO
Omic data expression analyses (genes, proteins, micro-RNA, exons)	YES	PARTIALLY

5. Discussion

As described by Terry et al.[3], there are five basic options for searching specific data in EHR: (i) pre-determined queries: users select a query option from the software

menu; (ii) simple customizable queries: users have some input into the queries to generate reports; (iii) advanced customizable queries: allow a greater amount of user input than the second level, often using Boolean logic; (iv) structured query language interface: using a special interface to enter Structured Query Language (SQL) commands; (v) data extraction and analysis with database tools. To date, the query language described in this paper is able to deal with levels 1 to 4 of Terry et al [3]. The global architecture of the underlying EHR system and the data querying strategy is closer to level 5 than to level 4 since, as reported by Terry et al [3] regarding level 5, the query language is based on the EHR's conceptual model. However, more advanced data analysis querying possibilities would probably be necessary to be considered as a full level 5 search options. Despite the fact the query language is quite complex to use, the public health professionals to whom it has been presented in fact stated that they would be able to use it after basic training. This training should also enable medical librarians, information scientists and IT specialists to use it. However, in contrast, several graphical user interfaces will be needed for health care professionals. These interfaces should provide access to more customizable queries than simple search. The I2B2 graphical interface could be a source of inspiration. To address this difficulty, an information extraction method was also designed in [4] to allow physicians to query EHRs using natural language instead of the dedicated QL. The SE has been tested outside the Rouen University Hospital, Normandy: at Bordeaux University Hospital, Aquitaine, France. However, the current model still does not operate on the establishment level but should become operational in the near future. Furthermore, the comparative evaluation of this query language with I2B2 should be improved. A parser enabling to share data between I2B2 data model and the RAVEL data model could be implemented to accurately assess precision as well as querying scope of the query languages. A scaling up study is underway at Rouen University Hospital with all the patients with at least one stay (in or outpatient) in the dermatology department since 1992 (n=65,000). This study aims at querying EHR data in a multi-patient context in order to create a patient cohort.

6. Acknowledgements

This search engine was partially funded by the French National Agency (TecSan program) in the RAVEL project.

References

[1] Prakash M Nadkarni. Qav: querying entity-attribute-value metadata in a biomedical database. Computer methods and programs in biomedicine, 53(2):93–103, 1997.

[2] Chloé Cabot, Julien Grosjean, Romain Lelong, Arnaud Lefebvre, Thierry Lecroq, Lina F. Soualmia, and Stéfan J. Darmoni. Omic data modelling for information on retrieval. In 2nd International Work-Conference on Bioinformatics and Biomedical Engineering, pages 415–424, 2014.

[3] Amanda L Terry, Vijaya Chevendra, Amardeep Thind, Moira Stewart, J Neil Marshall, and Sonny Cejic. Using your electronic medical record for research: a primer for avoiding pitfalls. Family Practice, 27(1):121–126, 2010.

[4] Lina F Soualmia, Romain Lelong, Badisse Dahamna, and St´efan J Darmoni. Rewriting natural language queries using patterns. In Multimodal Retrieval in the Medical Domain, pages 40–53. Springer, 2015.

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-126

Evaluation of the Terminology Coverage in the French Corpus LiSSa

Chloé CABOT[a], Lina F. SOUALMIA[a,b], Julien GROSJEAN[a], Nicolas GRIFFON[a,b] and Stéfan J. DARMONI[a,b,1]
[a] *Normandie Univ., TIBS - LITIS EA 4108, Rouen University and Hospital, France*
[b] *French National Institute for Health, INSERM, LIMICS UMR-1142, France*

Abstract. Extracting concepts from medical texts is a key to support many advanced applications in medical information retrieval. Entity recognition in French texts is moreover challenged by the availability of many resources originally developed for English texts. This paper proposes an evaluation of the terminology coverage in a corpus of 50,000 French articles extracted from the bibliographic database LiSSa. This corpus was automatically indexed with 32 health terminologies, published in French or translated. Then, the terminologies providing the best coverage of these documents were determined. The results show that major resources such as the NCI and SNOMED CT thesauri achieve the largest annotation of the corpus while specific French resources prove to be valuable assets.

Keywords. Information extraction, Semantics, Natural Language Processing, Data storage and retrieval, Vocabulary controlled

1. Introduction

Indexing medical documents such as clinical reports as well as biomedical articles is a key to various information retrieval tasks in medical information management. Automatic indexing can deal with the increasing amount of new material being produced in biomedical fields that has made manual indexing slow and expensive. Annotating medical documents and the following applications is actually a frequent topic in English-speaking scientific literature. Various annotating tools are available for English text, as well as the resources provided by the National Library of Medicine in association with the Unified Medical Language System (UMLS). Several vocabulary-controlled approaches for indexing documents have been proposed. Aronson et al. use MetaMap and the tri-gram method to extract UMLS terms, and then refine them to MeSH concepts [1]. Natural Language Processing (NLP) techniques can be also applied to annotate documents with UMLS [2]. Gurulingappa et al. use the JSRE system combining Support Vector Machines (SVMs) with different kernels specially designed for the NLP and relation extraction [3]. Vector space model (VSM) is also a common approach that can be mixed with NLP techniques. Jonnalagadda et al. adopt this approach to identify UMLS concepts in the i2b2/VA concept extraction corpus [4]

[1] Corresponding author, Stéfan J. DARMONI, Rouen University Hospital, 1 rue de Germont, 76000 Rouen, France; E-mail: stefan.darmoni@chu-rouen.fr

French-speaking texts do not benefit from such various tools and resources. French is lowly represented in the UMLS [5] As provided in the 2016AA release, the French UMLS thesaurus manages 9 resources while 128 resources are available in English, providing a French concept for 85,685 concept unique identifiers. Only 3.11% of English UMLS terms are available in French and while each English term has an average of 2 synonyms, only 1.54 synonyms are available for each French term.

Since 2005, our team develops the Health Terminology/Ontology Portal (HeTOP) [6] providing an access to 55 terminologies in French and English, partially translated into French. A major application of this multi-lingual portal includes a multi-terminology automatic indexing tool called ECMT [8] based on HeTOP resources.

The aim of this study is to analyze the coverage of 32 terminologies available in French in the HeTOP on the French medical corpus LiSSa [7]. These 32 terminologies were selected among the 55 available terminologies as the most relevant for this task. This corpus was indexed with the ECMT tool to help reduce (i) the amount of terminologies used in automatic indexing, (ii) the noise generated by using multiple terminologies, especially with some specific types of concepts and (iii) the amount of redundant concepts.

2. Methods

2.1. Automatic Indexing with ECMT

The ECMT tool is designed to identify clinical concepts in biomedical documents using terminologies included in HeTOP. ECMT relies on the "bag-of-words" algorithm and also on pattern-matching designed for discharge summaries, procedure reports or laboratory results which contain symbolic data (presence or absence), numerical data and units of measurement [8].

Each concept identified in a document and its metadata (the concept type, original identifier, terminology) was stored for subsequent analysis. Prior to the analysis of the coverage, the indexing terms, which presented the highest occurrence frequencies throughout the corpus, were manually reviewed to detect common and regular indexing errors, and excluded in relevant cases.

2.2. The French Medical Corpus LiSSa

The corpus of the bibliographic database LiSSa[2] contains more than 850,000 articles in French. Among them 50,000 articles were randomly selected and each title, abstract and set of keywords were indexed using the ECMT tool with 32 terminologies. These resources as well as the versions used are available in HeTOP[3]. The source language of these resources varies: 13 terminologies are published originally in French while 19 have been totally or partially translated.

[2] http://www.lissa.fr
[3] http://www.hetop.eu

3. Results

The amount of all concept occurrences identified in each terminology is determined for each document category: titles, abstracts and sets of keywords. The results are detailed in Table 1 and Figure 1. Distinct concepts (i.e. counted only once)

Table 1. Terminology coverage of the French corpus LiSSa for each document category.

Titles		Abstracts		Keywords	
Terminology	**Concepts**	**Terminology**	**Concepts**	**Terminology**	**Concepts**
NCIt	150,224	NCIt	2,040,356	NCIt	52,423
MeSH	106,170	SNOMED CT	1,543,456	MeSH	50,937
SNOMED Int.	96,771	MeSH	1,238,133	TSP	47,237
SNOMED CT	95,409	TSP	1,089,331	SNOMED Int.	45,879
TSP	84,989	SNOMED Int.	827,714	SNOMED CT	36,771
MedDRA	45,164	LOINC	502,964	MedDRA	25,802
LOINC	36,483	MedDRA	395,434	LOINC	15,491
FMA	24,300	FMA	182,398	ICNP	13,585
ICNP	24,244	ICNP	161,341	FMA	8,819
ICD-10	14,022	CLADIMED	87,924	HPO	7,633
Others	77,273	Others	641,766	Others	40,583
Total	755,049	Total	8,710,817	Total	345,160

identified in each terminology were also determined for each document category. The results are detailed in Table 2.

The five terminologies obtaining the most indexing terms in each document category, NCIt, SNOMED CT, SNOMED Int., MeSH and TSP are consistently the same for each group. The NCI thesaurus obtains the best coverage in all document categories, while the Thésaurus Santé Publique (TSP), a French Public Health thesaurus is the only French ressource to appear in the first third of the ressource ranking. More specialized resources such as HRDO (rare diseases) or ATC (chemical therapeutics) achieve much less coverage than expected. The five first terminologies giving the best coverage of the corpus add up 65% to 70% of the whole indexing term set depending on the document category. However, some smaller resources published originally in French achieve a good coverage of the corpus in spite of a limited French indexing terms. These resources such as the CISMeF thesaurus [9] or the Q-Codes classification [10] are actually developped to fit clinical and non-clinical information in abstracts and complete larger terminologies as the MeSH or the SNOMED Int.

Table 2. Terminology coverage by distinct concepts of the French corpus LiSSa for each document category.

Titles		Abstracts		Keywords	
Terminology	**Concepts**	**Terminology**	**Concepts**	**Terminology**	**Concepts**
MeSH	11,324	MedDRA	19,281	MeSH	5,908
SNOMED Int.	10,396	SNOMED Int.	17,968	SNOMED Int.	4,290
MedDRA	8,644	MeSH	17,353	NCIt	4,249
SNOMED CT	7,996	SNOMED CT	17,192	MedDRA	4,024
NCIt	7,247	NCIt	12,683	SNOMED CT	3,491
TSP	3,617	TSP	5,799	TSP	2,801
LOINC	1,822	FMA	3,903	LOINC	1,012
FMA	1,815	LOINC	3,372	FMA	935
ICD-10	1,758	HPO	2,997	ICD-10	870
HPO	1,488	ICD-10	2,812	HPO	823
Others	7,186	Others	11,612	Others	4,082
Total	63,293	Total	114,982	Total	32,485

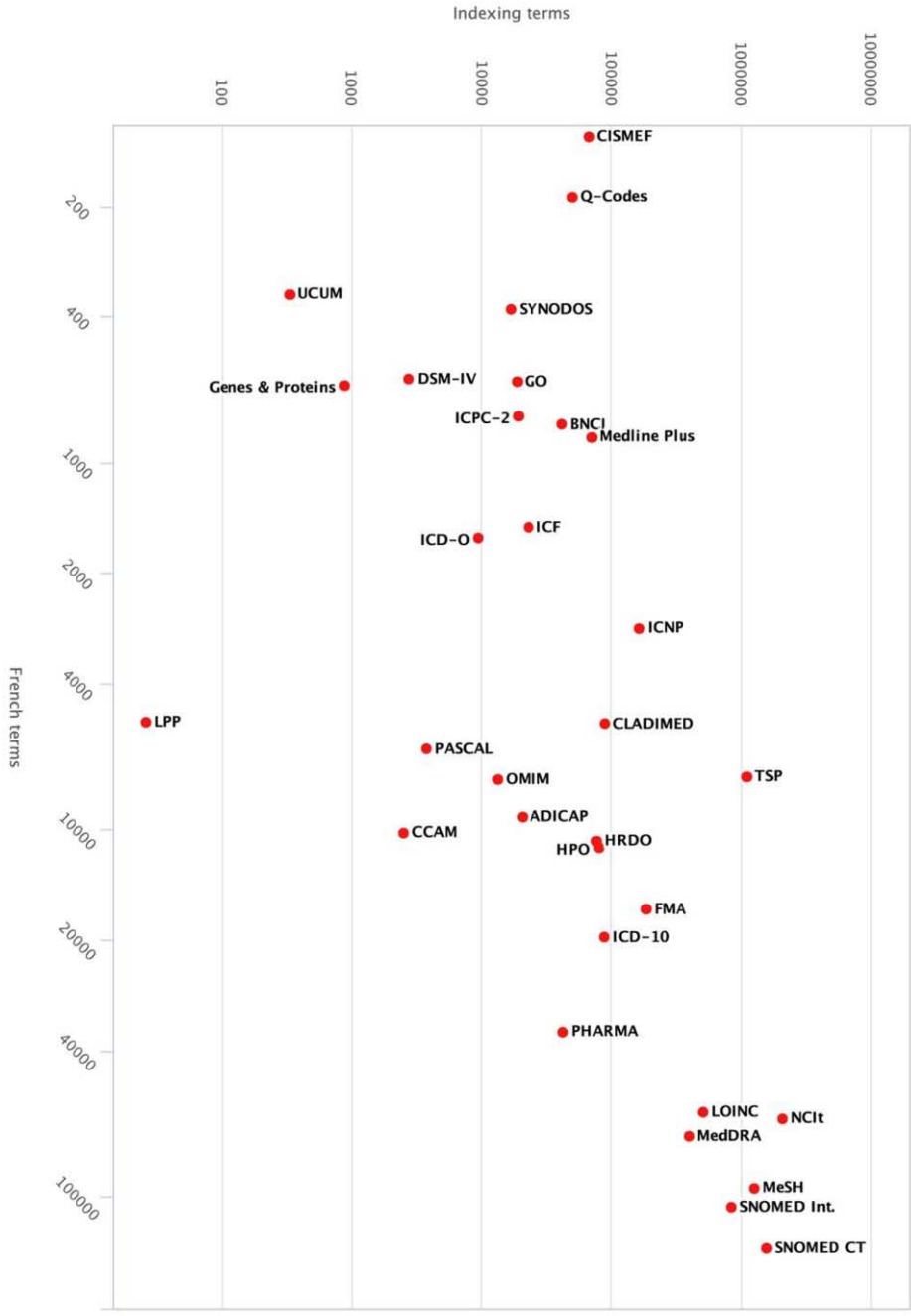


Figure 1. Terminology coverage of article abstracts in the LiSSa corpus.

4. Discussion and Conclusion

LiSSa is a bibliographic database in French providing a large corpus of titles, abstracts and authors' keywords. ECMT was able to annotate these three corpora. Overall, NCIt is surprisingly ranked first in the three corpora although only 60000 terms are now translated in French while over 90000 are translated for the MeSH and 137000 for SNOMED CT. When analyzing with distinct concepts, the ranking is very different (MeSH ranked first for titles, MedDRA for abstracts). This coverage of distinct concepts should be refined at the concept scale to evaluate the concept redundancy between the top ten ranked resources. This study is still ongoing and a phase of manual annotation of the corpus by field experts to validate the automatic indexing results and refine these observations is currently processed. In the near future, we will reproduce the same study on a corpus of discharge summaries. Terminologies developed for this purpose should be better ranked, in particular SNOMED CT and ICD-10.

5. Acknowledgments

The LiSSa project was partially granted by the ANR TecSan program (ANR-14-CE17-0020).

References

[1] Alan R Aronson, James G Mork, Clifford W Gay, Susanne M Humphrey, and Willie J Rogers. The NLM Indexing Initiative's Medical Text Indexer. *Studies in health technology and informatics*, 107(Pt 1):268–272, 2004.

[2] Asma Ben Abacha and Pierre Zweigenbaum. Automatic extraction of semantic relations between medical entities: a rule based approach. *Journal of biomedical semantics*, 2 Suppl 5(Suppl 5):S4, October 2011.

[3] Harsha Gurulingappa, Abdul Mateen-Rajput, and Luca Toldo. Extraction of potential adverse drug events from medical case reports. *Journal of biomedical semantics*, 3(1):15, December 2012.

[4] Siddhartha Jonnalagadda, Trevor Cohen, Stephen Wu, and Graciela Gonzalez. Enhancing clinical concept extraction with distributional semantics. *Journal of biomedical informatics*, 45(1):129–140, February 2012.

[5] A Névéol, J Grosjean, S J Darmoni, and P Zweigenbaum. Language Resources for French in the Biomedical Domain. *LREC*, 2014.

[6] J Grosjean, T Merabti, and B Dahamna. Health multi-terminology portal: a semantic added-value for patient safety. *Stud Health Technol*, 2011.

[7] Nicolas Griffon, Matthieu Schuers, Lina Fatima Soualmia, Julien Grosjean, Gaetan Kerdelhué, Ivan Kergourlay, Badisse Dahamna, and Stéfan Jacques Darmoni. A Search Engine to Access PubMed Monolingual Subsets: Proof of Concept and Evaluation in French. *Journal of Medical Internet Research*, 16(12):e271, December 2014.

[8] L F Soualmia, C Cabot, B Dahamna, and S J Darmoni. SIBM at CLEF e-Health Evaluation Lab 2015. 2015.

[9] Magaly Douyère, Lina F Soualmia, Aurelie Neveol, Alexandrina Rogozan, Badisse Dahamna, Jean-Philippe Leroy, Benoit Thirion, and Stéfan J Darmoni. Enhancing the MeSH thesaurus to retrieve French online health resources in a quality-controlled gateway. *Health Information & Libraries Journal*, 21(4):253–261, 2004.

[10] M Jamoulle. Using the International Classification for Primary Care (ICPC) and the Core Content Classification for General Practice (3CGP) to classify conference abstracts. *Rev Port Med Geral Fam*, 2013.

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-131

Linked Data Applications Through Ontology Based Data Access in Clinical Research

Ann-Kristin KOCK-SCHOPPENHAUER [a], Christian KAMANN [a, b], Hannes ULRICH [a, b], Petra DUHM-HARBECK [a], Josef INGENERF [a, b, 1]
[a] *IT for Clinical Research, Lübeck (ITCR-L), University of Lübeck, Germany*
[b] *Institute of Medical Informatics, University of Lübeck, Germany*

Abstract. Clinical care and research data are widely dispersed in isolated systems based on heterogeneous data models. Biomedicine predominantly makes use of connected datasets based on the Semantic Web paradigm. Initiatives like Bio2RDF created Resource Description Framework (RDF) versions of Omics resources, enabling sophisticated Linked Data applications. In contrast, electronic healthcare records (EHR) data are generated and processed in diverse clinical subsystems within hospital information systems (HIS). Usually, each of them utilizes a relational database system with a different proprietary schema. Semantic integration and access to the data is hardly possible. This paper describes ways of using Ontology Based Data Access (OBDA) for bridging the semantic gap between existing raw data and user-oriented views supported by ontology-based queries. Based on mappings between entities of data schemas and ontologies data can be made available as materialized or virtualized RDF triples ready for querying and processing. Our experiments based on CentraXX for biobank and study management demonstrate the advantages of abstracting away from low level details and semantic mediation. Furthermore, it becomes clear that using a professional platform for Linked Data applications is recommended due to the inherent complexity, the inconvenience to confront end users with SPARQL, and scalability and performance issues.

Keywords. Semantic Web, Linked Data, Semantic Querying and Data Integration

1. Introduction

Within translational research there is a demand to semantically process and integrate clinical care and biomedical research data from different heterogeneous resources. Instead of schema matching approaches (e.g. for relational databases) the Semantic Web paradigm uses the Resource Description Format (RDF) for the flexible representation of facts together with a semantic layer for describing corresponding types and relationships by ontologies (RDFS, OWL). Efficient frameworks for distributed queries across multiple RDF data sources are used in many application areas; amongst others in biomedicine and less often in healthcare [2, 3]. In clinical care usually Relational Database Management Systems (RDBMS), i.e. mature products like Oracle, MS SQL or MySQL, are used.

Ontology-Based Data Access (OBDA) is a new paradigm for accessing and integrating data, whose key concept is to resort to a three-level architecture with an ontology,

[1] Corresponding author, Josef Ingenerf, Institut für Medizinische Informatik, Ratzeburgerallee 160, 23562 Lübeck, Germany; E-mail: ingenerf@imi.uni-luebeck.de.

data sources, and mappings between both [4]. The ontology defines a high-level global schema of data sources and provides a vocabulary in terms of concepts, roles, i.e. binary relations and attributes for user queries. The mapping layer explicitly specifies the relationships between the domain concepts and the data sources. Afterwards an OBDA system rewrites such queries and ontologies into the vocabulary of the data sources and delegates the actual query evaluation to a suitable query answering system such as SQL for RDBMS. The ontology together with the mappings exposes a virtual RDF graph, which can be queried using SPARQL, the standard query language for RDF data. This virtual RDF graph can be materialized by using RDF triplestores, or alternatively it can be kept virtual and queried only during query execution.

2. Methods and Material

In the following we present several ways to adopt the OBDA approach by accessing data from the RDBMS-based CentraXX system for biobank and study management [5].

2.1. Ontop Used As a Plugin within Protégé

Ontop is one of the most popular OBDA systems [6]. This open source software is available amongst others as plugin for the ontology editor Protégé. First, a domain ontology with relevant concepts like patients, encounters and diagnoses and their relationships are defined. Second, original data sources are connected and mappings are managed. The mapping includes how classes of instances and relationships are mapped to the database entries by SQL statements, see Fig 1. Finally, SPARQL queries are created and executed by *Quest*, a query answering engine with OWL 2 QL/RDFS entailment.

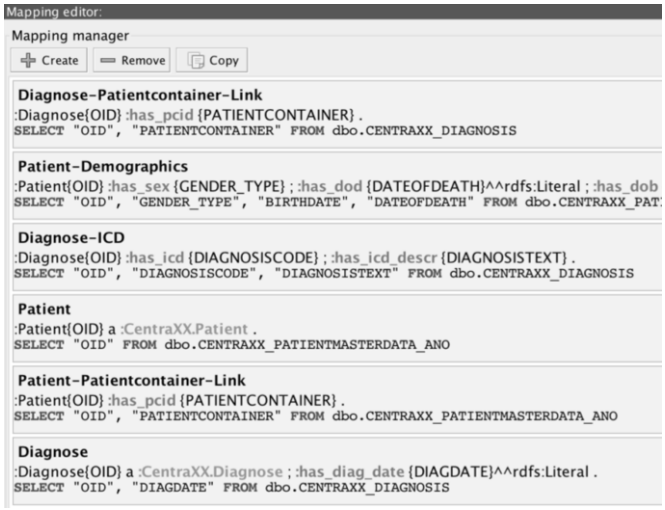


Figure 1. Realizing OBDA with the Ontop plugin within Protégé, applied to the CentraXX database.

The mappings expressed in W3C standard R2RML (RDB to RDF Mapping) can be constructed semi-automatically by domain experts or by using a bootstrapper that creates the ontology and mappings automatically by analyzing the database schema. Furthermore, Ontop works with Teiid as open source Java software for data virtualization, used

for federating different heterogeneous RDBMS behind one JDBC interface. The use of Ontop as plugin in Protégé was not fully sufficient, for mainly two reasons: First, the end users of the envisioned OBDA-based query system should not be forced to enter SPARQL queries. Second, there is a limitation when trying to follow up with linked data applications based on the resulting RDF triples.

2.2. Optique Platform with OptiqueVQS as a Visual Query System

Optique (Scalable End-user Access to Big Data) is an EU-funded project where novel solutions based on the OBDA idea have been developed [7]. The Optique Platform has been made available as an app that can be installed and deployed within the Information Workbench, see chapter 2.3. The platform features a visual query system (VQS) where query dialogues are rendered based on the ontology, see Fig 2. However, in spite of the impressive potential of the VQS, the flexibility with regard to the desired query frontend was not sufficient, since this can mainly be influenced by cumbersome modifications of the ontology.

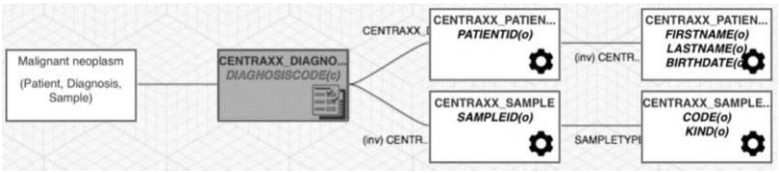


Figure 2. Screenshot of the query interface generated by OptiqueVQS.

2.3. Information Workbench (IWB) - A Platform for Linked Data Applications

For flexibility reasons we decided to work directly with the Information Workbench (IWB) [8]. The software provides a generic frontend for customizable user interfaces based on Semantic Wiki technologies, enriched with a large set of widgets for data ac-

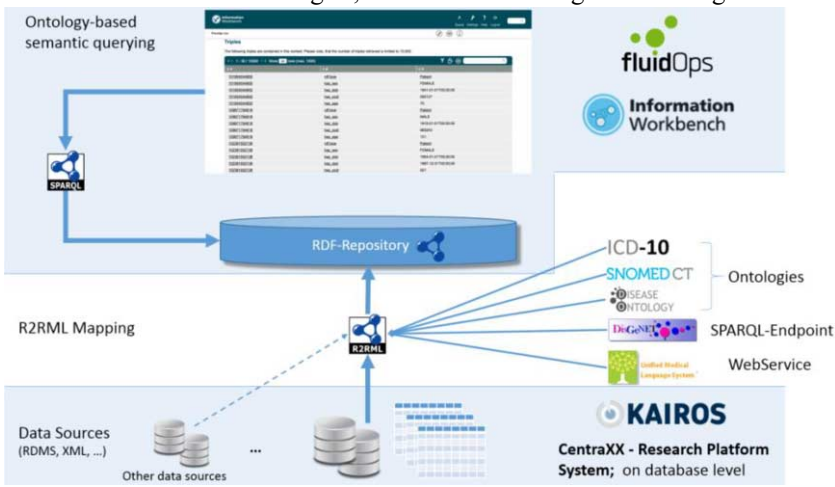


Figure 3. Information Workbench for linked data applications based on relational data from CentraXX.

cess, navigation, visualization, analytics and data mashups with external data sources [9]. It can be further customized and extended for domain specific applications through a SDK. Techniques for OBDA from chapter 2.1 and 2.2 like Ontop are included in the platform.

For federated queries, i.e. sending decomposed SPARQL subqueries to various data services and integrating the results virtually, the FedX module is available [10]. The Information Workbench is available as a Community Edition under an open source license as well as an Enterprise Edition with a commercial license. We used IWB for creating a demonstrator that provides an ontology based query frontend for accessing RDBMS data of the CentraXX system enriched by linked data, see Fig 3.

3. Results and Discussion

The IWB greatly facilitates the creation of an ontology with concepts, relationships and attributes of interest and relevant R2RML-mappings to the relational database used for example by CentraXX. This allows to materialize corresponding RDF triples internally.

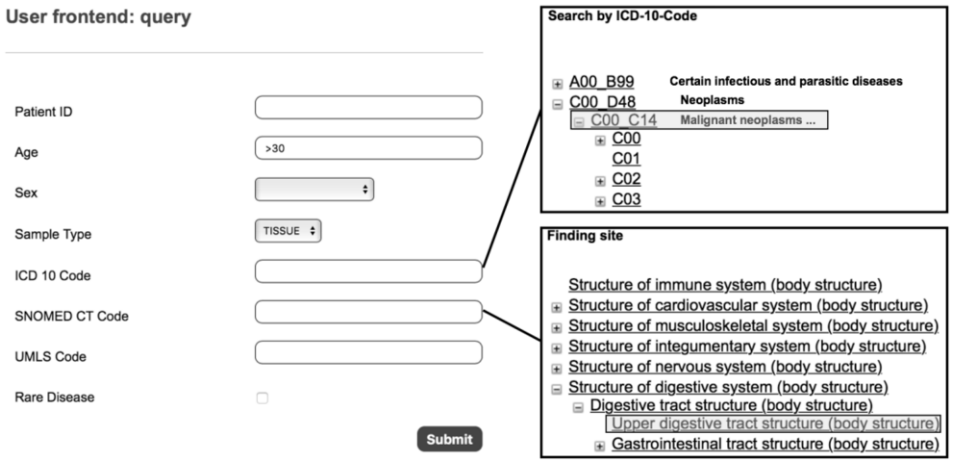


Figure 4. User frontend and technical details of the demonstrator using the IWB

By using further ontologies like the Disease Ontology [11] or the Unified Medical Language System (UMLS) [12] the triples could be semantically enriched by ontological mappings, especially taking the ICD-10 codes within the CentraXX data into consideration. By extracting finding sites from mapped disorder concepts in SNOMED CT and adding this anatomical codes to the RDF repository. For example, it becomes possible to query all data from patients or samples that are located at the digestive tract. An example for such a query is shown in Fig 4. Compared to ICD-10 based retrieval this is an example for the kind of added values that should be further explored. Additionally, it is possible to use annotated Disease Ontology or UMLS codes for accessing linked data of interest like disease associated genes from the SPARQL endpoint of DisGenNET [13]. There are much more linked data of interest that we might include similarly, e.g. accessing literature from MEDLINE.

On the userinterface, the customization is done in a completely declarative way, resorting to a rich pool of widgets and creating template pages in wiki syntax, which are associated with elements of domain ontologies. In our preliminary experiments this significantly simplifies and speeds up the application development. The OBDA paradigm is a promising approach within clinical research informatics because a lot of existing applications are based on relational database systems.

Acknowledgement

The authors would like to thank fluid Operations AG and KAIROS GmbH for their support. The approach presented in this paper is implemented on a CentraXX test environment, but can be easily transferred to the system in use as well as to many other EHR systems of interest.

References

[1] Belleau F, Nolin MA, Tourigny N, Rigault P, Morissette J: *Bio2RDF: towards a mashup to build bioinformatics knowledge systems*. J Biomed Inform. 2008; 41(5):706-16.

[2] Carmen Legaz-Garcia MD, Minarro-Gimenez JA, Menarguez-Tortosa M, Fernandez-Breis JT: *Generation of open biomedical datasets through ontology-driven transformation and integration processes*. J Biomed Semantics. 2016; 7:32. 10.1186/s13326-016-0075-z.

[3] Hussain S, Ouagne D, Sadou E, Dart T, Jaulent MC, De Vloed B, Colaert D, Daniel C: *EHR4CR: A Semantic Web Based Interoperability Approach for Reusing Electronic Healthcare Records in Protocol Feasibility Studies*. In: Proceedings of the 5th International Workshop on Semantic Web Applications and Tools for Life Sciences; 2012. p. http://ceur-ws.org/Vol-952/paper_31.pdf.

[4] Liaw ST, Taggart J, Yu H, de Lusignan S, Kuziemsky C, Hayen A: *Integrating electronic health record information to support integrated care: practical application of ontologies to improve the accuracy of diabetes disease registers*. J Biomed Inform. 2014; 52:364-72. 10.1016/j.jbi.2014.07.016

[5] CentraXX - KAIROS GmbH, [Internet, cited 19 October 2016], Available from: http://www.kairos.de/centraxx/.

[6] Calvanese D, Cogrel B, Komla-Ebri S, Kontchakov R, Lanti D, Rezk M, Rodriguez-Muro M, Xiao G: *Ontop: Answering SPARQL queries over relational databases*. Semantic Web. 2016:1-17 [in press].

[7] Giese M, Soylu A, Vega-Gorgojo G, Waaler A, Haase P, Jiménez-Ruiz E, Lanti D, Rezk M, Xiao G, Özccep ÖL, Rosati R: *Optique: Zooming in on Big Data*. IEEE Computer. 2015; 48(3):60-67.

[8] Information Workbench - Fluid Operations AG, [Internet, cited 19 October 2016], Available from: https://www.fluidops.com/en/products/information_workbench/.

[9] Gossen A, Haase P, Hütter C, Meier M, Nikolov A, Pinkel C, Schmidt M, Schwarte A: *The Information Workbench - A Platform for Linked Data Applications*. Semantic Web 2016:1-7 [in press].

[10] Schwarte A, Haase P, Hose K, Schenkel R, Schmidt M: *FedX: Optimization Techniques for Federated Query Processing on Linked Data*. In: The Semantic Web - ISWC 2011 - 10th Internat. Semantic Web Conference, Bonn, Germany, October 23-27, 2011, Proceedings, Part I: Springer; 2011. p. 601-616.

[11] Schriml LM, Arze C, Nadendla S, Chang YW, Mazaitis M, Felix V, Feng G, Kibbe WA: *Disease Ontology: a backbone for disease semantic integration*. Nucleic Acids Res. 2012; 40 (Database issue): D940-6. 10.1093/nar/gkr972

[12] Bodenreider O: *Biomedical ontologies in action: role in knowledge management, data integration and decision support*. Yearb Med Inform. 2008:67-79.

[13] Queralt-Rosinach N, Pinero J, Bravo A, Sanz F, Furlong LI: *DisGeNET-RDF: harnessing the innovative power of the Semantic Web to explore the genetic basis of diseases*. Bioinformatics. 2016; 32(14):2236-8. 10.1093/bioinformatics/btw214

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-136

Epidemiological Models Lacking Process Noise Can Be Overconfident

Lavi SHPIGELMAN[a1], Michal CHOREV[a], Zeev WAKS[a], Ya'ara GOLDSCHMIDT[a], Edwin MICHAEL[b]
[a] *IBM Research Lab, Haifa, Israel*
[b] *University of Notre Dame, IN, USA*

Abstract. Mathematic models of epidemics are the key tool for predicting future course of disease in a population and analyzing the effects of possible intervention policies. Typically, models that produce *deterministic* are applied for making predictions and reaching decisions. Stochastic modeling methods present an alternative. Here, we demonstrate by example why it is important that *stochastic* modeling be used in population health decision support systems.

Keywords. Epidemiological models, Stochastic processes

1. Introduction

Epidemiological processes are usually derived from probabilities of disease events. In spite of this, arguably, most epidemiological *modeling* studies employ deterministic dynamical systems, usually (Ordinary) Differential Equations (ODEs) to simulate the processes and make predictions. Classic textbooks such as [1] do not elaborate on stochastic approaches. Apparently, despite the more true-to-reality nature of stochastic models, the value of stochasticity is presumed to be low.

All models are wrong by some measures [2]. For an epidemiological model to be useful (as they often are in decision support and control frameworks), it needs to be accurate, transparent (its *logic* easily understood, if not its *dynamics*) and flexible [3]. As Andersson and Britton [4] claim, "stochastic models are to be preferred when their analysis is possible". This is because typically they do well by all 3 criteria.

There are several ways of introducing stochasticity into models that may produce benefits and they stem from different reasons. One is **parameter uncertainty:** Since model parameters are *unknown* exactly, sampling them from (posterior) distributions and using the sampled values to propagate simulated model trajectories would reflect that uncertainty. Another is **event driven process stochasticity:** when the target population is small or the number of individuals in a group (e.g. the number of infected) is small, fluctuations due to the randomness of transitions of individuals is significant. A third reason is **stochasticity due to inaccurately modeled processes:** By their nature, models typically leave out multiple time-varying processes that affect the modeled variables. Unless a process is the specific focus of the study or its effects are critical for reasonably accurate modeling, it is typically left out. The effects of

[1] Corresponding author, IBM Research Lab, Haifa University Campus, Haifa, 3498825, Israel. Email: Lavi@il.ibm.com

unmodeled processes need not diminish as population sizes increase and they can add up to significant process "noise". Last but not least is **observation noise:** Most observational data in epidemiology is the result of partial, biased, error prone sampling of the real system. Therefore, observations are not exact measures of the model variables. The difference can be accounted for by modeling the observations as a distribution that is conditioned on the model state variables.

Introducing probabilistic aspects to models does, however, come at a price in terms of model developer time, computation time, number of state variables or the complexity of their interactions. In recent years, however, the availability of cheap computation-intensive hardware and the development of general purpose probabilistic programming tools [5] such as Stan [6] and PyMC [7] have substantially lowered the difficulty of such tasks. Unfortunately, the literature on the subject is typically not aimed at epidemiologists or is math-heavy and technical. Additionally, the literature often focuses on the analysis of the dynamics of such models rather than on *estimating their parameters* (e.g. [8, 3]). The goal of this paper is to provide a gentle exposition by example. We present the deterministic (ODE based) and re-stochastised versions of a basic epidemiological model (The SIRS compartmental model) and learn its parameters using Stan. We then analyze the learned models and show that estimating parameter distributions and simulating predicted trajectories with the deterministic ODE model (allowing only for parameter uncertainty) can lead to significant inaccuracies in expected values and confidence ranges compared with its stochastic counterpart.

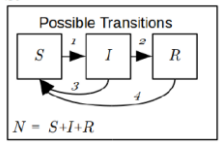
2. Methods

Simulations and parameter estimation were performed with code written in Stan [6, 9]. Stan is a freely available probabilistic programming language for specifying statistical models and a tool for sampling from those models, conditioned on observations, using state-of-the-art self -tuning MCMC methods.

a.

	event i	$d\boldsymbol{x}_i =$ $d(S,I,R)$	rate $r(d\boldsymbol{x}_i)$
1	infection	-1, 1, 0	$\beta \frac{I}{N} S$
2	recovery with immunity	0, -1, 1	γI
3	recovery without immunity	1, -1, 0	bI
4	loss of immunity	1, 0, -1	bR

b.



c.

$$\boldsymbol{\mu} = \begin{bmatrix} \frac{dS}{dt} \\ \frac{dI}{dt} \\ \frac{dR}{dt} \end{bmatrix} = \overbrace{\begin{bmatrix} -\frac{\beta SI}{N} + b(I+R) \\ \frac{\beta SI}{N} - (b+\gamma)I \\ \gamma I - bR \end{bmatrix}}^{\sum r(d\boldsymbol{x}_i)\, d\boldsymbol{x}_i} \quad \boldsymbol{G} = \begin{bmatrix} -1 & 0 & 1 & 1 \\ 1 & -1 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{bmatrix} \cdot \begin{bmatrix} \sqrt{\beta \frac{I}{N} S} & 0 & 0 & 0 \\ 0 & \sqrt{\gamma I} & 0 & 0 \\ 0 & 0 & \sqrt{bI} & 0 \\ 0 & 0 & 0 & \sqrt{bR} \end{bmatrix} = \begin{bmatrix} -\sqrt{\beta \frac{I}{N} S} & 0 & \sqrt{bI} & \sqrt{bR} \\ \sqrt{\beta \frac{I}{N} S} & -\sqrt{\gamma I} & -\sqrt{bI} & 0 \\ 0 & \sqrt{\gamma I} & 0 & -\sqrt{bR} \end{bmatrix}$$

$$d\boldsymbol{x}(t) = \boldsymbol{\mu}(\boldsymbol{x}(t))\, dt + \boldsymbol{G}(\boldsymbol{x}(t))\, d\mathbf{W}(t) \qquad (1)$$

Figure 1 SIR (Susceptible - Infected - Recovered) model. **a:** Table of possible events. **b.** Graphical representation of compartments and transitions. **c.** Definition of $\boldsymbol{\mu}(\boldsymbol{x})$, the deterministic ODE model and of **G**, the covariance matrix (based on the events table). Eq. (1) is the resulting SDE model.

2.1. Deriving Dynamic Bayesian Models (DBNs) from ODEs and Model Simulations

Most compartmental models are described as ordinary differential equations (ODEs), $d\boldsymbol{x}/dt = \boldsymbol{\mu}(\boldsymbol{x}(t))$, $\boldsymbol{x} \in \mathbb{R}^d$, $\boldsymbol{\mu}: \mathbb{R}^d \to \mathbb{R}^d$ where d is the number of compartments or state variables. These equations in turn stem from a set of possible independent transition events from one compartment to another or injections (a.k.a.

births) that are sometimes called the master equation description of a system. Allen [8] describes a Standard method for converting such ODE systems into Stochastic Differential Equations (SDEs), whose Euler integration at (sufficiently small) fixed time interval comprises a Standard Dynamic Bayesian Model (DBN) where the transitions from time t to $t + dt$ are normally distributed with $\boldsymbol{\mu}(\boldsymbol{x}(t))$ as their mean and covariance, $\boldsymbol{V}(\boldsymbol{x}(t)) = \boldsymbol{G}\boldsymbol{G}^T$ where $\boldsymbol{G}(\boldsymbol{x}(t))$ is a matrix function of the previous state derived from the events and their state dependent rates. For the sake of brevity, we explain this derivation method by example for the classic SIR model in Figure 1 (with fixed population size and loss of immunity, a.k.a. SIRS), which often serves as a basis for more specialized models. This model is typically implemented as the deterministic ODE system $\boldsymbol{\mu}(\boldsymbol{x})$ of Figure 1.c . The (Euler integration of the) SDE model is given in differential form in Eq. (1). $d\mathbf{W}(t)$ is a Wiener process, which, in this Euler integration is simply a Standard normally distributed noise variable $N(0, I) \in \mathbb{R}^4$ (dimension corresponding to the number of possible event types) multiplied by $\sqrt{dt}$, hence describing a DBN with non linear Gaussian transitions. Note that other approaches could be taken for converting an ODE model to a DBN. In [10], a direct approach is taken, letting modelers choose the types of distributions as they see fit. They also describe how to perform more accurate integration steps such as Runge Kutta.

We chose to present the SDE method here because the introduction of stochasticity is principled and stems from the original assumed sources of event uncertainty. We further modified the SDE in some of the simulations to model a parameter which is not constant but is, rather, fluctuating with time. Specifically, we assumed that the infection force, β, is an Ornstein–Uhlenbeck process [11]. I.e. its course is described by $d\beta(t) = \theta(\mu\beta - \beta t)dt + \sigma\, dWt$ where $\boldsymbol{\mu_\beta}$ is the equilibrium mean of the fluctuations (set to 1.0) , σ (0.05) is a volatility parameter and θ (also 0.05) is a dissipation rate. Thus $\beta(t)$ in this scenario acts as a fourth state parameter whose integrated dynamics are determined by the independent Gaussian samples. Details of the simulations performed are in Figure 2. Since these simulations were not conditioned on observations, the sampling did not involve Stan's MCMC sampling engine (just simple sampling from prior distributions)

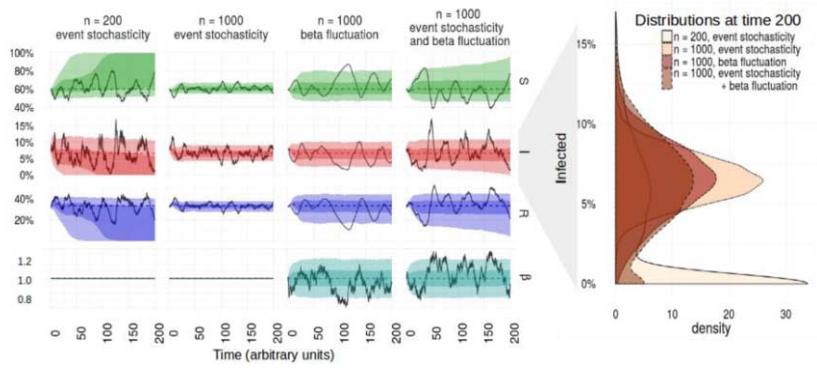


Figure 2 SIR Simulations. **a**: simulation results for S, I and R populations (as % of total) and the β variable. Columns differ in total population size and sources of noise. Solid lines represent single sample trajectories (one per column). Dashed lines are the deterministic ODE trajectories without noise. Ribbons show the 5th, 25, 75, 95 percentiles for 20,000 simulations. **b**: Estimated distributions (smoothed histograms) of the infected population at the last time point for all the configurations (columns on left). All simulations were performed for 200 time units at $dt = 0.1$ with parameters $\beta = 1.0$ (+ Wiener process fluctuation), $\gamma = 0.5$ and $b = 0.1$ from fixed initial conditions close to equilibrium (S=60%, I= 8%, R= 32%)

2.2. Parameter Fitting

Parameter fitting for β, b and γ was performed in a semi-realistic scenario: A single trajectory for a population of size $N = 1000$ was generated with both β and event-driven stochasticity. Observations were generated at 20 time points along the trajectory by sampling the number of infected versus not-infected from a binomial distribution (assuming sampling with replacement) with sample size of 200 and rate determined by the (hidden) true proportion of infected individuals in the population. Parameter learning (by posterior sampling) used Stan's NUTS MCMC method [12] (with 4 chains and 2000 iterations, 1000 burn-in, to produce 4000 samples) with uniform parameter priors (parameter stochasticity) in the ranges shown in Figure (3.c), from Dirichlet distributed initial starting points (Dirichlet α parameter at $10\times$ the true initial proportions). Fitting was performed twice: using either the ODE model (with parameter uncertainty but no process noise) or the DBN model (SDE implementation as described by Eq. (1) with β fluctuations with known μ_β and θ parameters).

3. Results

Figure (2) shows S, I, R and β trajectories simulated with fixed known parameters and the distributions at the last simulation time point comparing different conditions. As expected, a smaller population size leads to an increase in process noise due to the

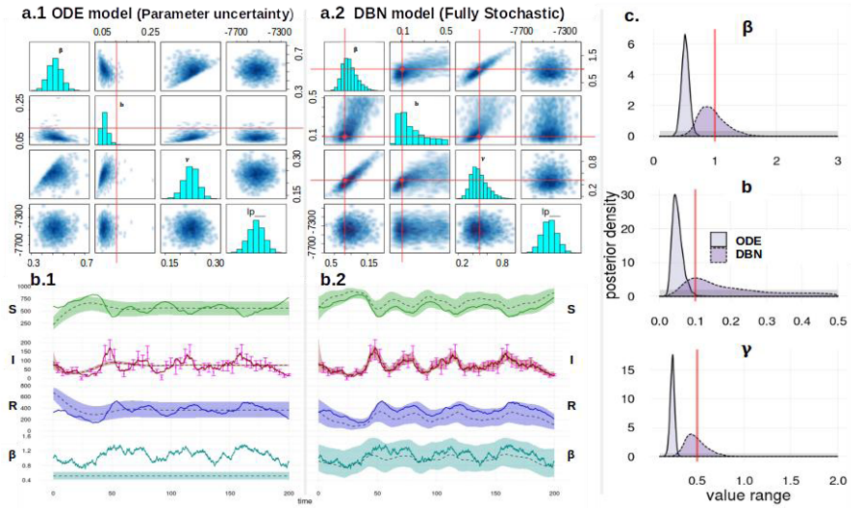


Figure 3 Parameter fitting for ODE model (left) and SDE model (center). **a:** Pairs scatter plot of posterior samples of the parameters and log probability ("lp"). Red lines indicate true values when within range. **b:** Posterior sampled trajectory distributions for S, I, R and β (not time dependent on left). Dashed lines are the median trajectories. Ribbons are 90% empirical confidence intervals (between 5th and 95th percentiles). Solid lines are the true (unobserved) values. Magenta points + bars are the sampled observations of infecteds (scaled appropriately) and their 2 Standard-deviation confidence intervals (binomial with sample size of 200). **c:** Single parameter histograms overlaid for the two models (corresponding to the pairs plots' diagonals). Plot ranges reflects the prior distribution ranges.

uncertainty of the events driving the process. The introduction of fluctuations in the β variable add noise whose variance does not diminish with population size (variance of $\beta(t)$ at equilibrium around μ_β is $\sigma^2/2\theta$). Importantly, even for a large population, the effects of event-caused variability may be significant when one of the compartments is

small, as is the case for the infected population in this simulation. Furthermore, such fluctuations may even move the system from the diseased equilibrium to the disease extinction state with non-negligible probability.

The model learning (posterior sampling) results are shown in Figure 3. It is clear that the fully stochastic model was better able to fit the fluctuating signals while also producing parameter distributions that better cover the parameters of the true trajectory. The model employing only parameter stochasticity produced too narrow posterior parameter distributions, giving the true parameters essentially zero probability and producing trajectory distributions that fail to capture not only the fluctuations themselves (as would be expected) but also their ranges, especially of I, resulting in trajectories that do not even overlap with many of the observations' confidence intervals. Such a model might still make sensible predictions of expected equilibria and short term expected median behavior but would be bad at estimating the correct confidence intervals which are essential for answering queries regarding robustness of predictions and chances of extreme events.

4. Summary

We showed here an analyzed example of the effects of modeling process noise on epidemiological process analysis and prediction. Our contribution is not to the models and algorithms, but rather to the analysis of the effects of properly modeling process noise on parameter estimation and trajectory fitting. We show that by limiting the modeled sources of uncertainty there is a real risk of deriving overly narrow confidence intervals for parameters and trajectories even when working within the Bayesian framework of parameter estimation. This may lead to ineffective policies when such models are used in decision making regarding disease intervention and control. Given that currently existing tools are advanced enough for stochastic process dynamics, modelers should consider using them to improve parameter estimations and predictions.

References

[1] K. J. Rothman, S. Greenland and T. L. Lash, Modern epidemiology, Williams & Wilkins, 2008.
[2] G. E. P. Box, Science and Statistics, J. Am. Statistical Association, vol. 71, no. 356, pp. 791-799, 1976.
[3] M. J. Keeling and P. Rohani, Modeling infectious diseases in humans and animals, Princeton University Press, 2008.
[4] H. Andersson and T. Britton, Stochastic epidemic models and their statistical analysis, Springer, 2012.
[5] A. D. Gordon, T. A. Henzinger, A. V. Nori and S. K. Rajamani, "Probabilistic programming," in Proceedings of the on Future of Software Engineering, 2014.
[6] B. Carpenter, Stan: A Probabilistic Programming Language, Journal of Statistical Software, 2015.
[7] Patil, D. Huard and C. J. Fonnesbeck, PyMC: Bayesian stochastic modelling in Python, Journal of statistical software, vol. 35, no. 4, p. 1, 2010.
[8] E. Allen, Modeling with Ito stochastic differential equations, vol. 22, Springer, 2007.
[9] Stan Modeling Language User's Guide and Reference Manual, Version 2.10.0, 2015.
[10] C. G. Enright, et al, Bayesian networks for mathematical models: Techniques for automatic construction and efficient inference, Int. J. Approximate Reasoning, vol. 54, no. 2, pp. 323-342, 2013.
[11] J. L. Doob, The Brownian movement and stochastic equations, Ann. Mathematics, pp. 351-369, 1942.
[12] M. D. Hoffman and A. Gelman, The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo, J. Machine Learning Research, 2014.
[13] D. Koller and N. Friedman, Probabilistic Graphical Models: Principles and Techniques - Adaptive Computation and Machine Learning, The MIT Press, 2009.

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-141

Disentangling Prognostic and Predictive Biomarkers Through Mutual Information

Konstantinos SECHIDIS [a,1], Emily TURNER [a], Paul METCALFE [b], James WEATHERALL[b] and Gavin BROWN [a]
[a] *School of Computer Science, University of Manchester*
[b] *Advanced Analytics Centre, Global Medicines Development, AstraZeneca*

Abstract. We study information theoretic methods for ranking biomarkers. In clinical trials, there are two, closely related, types of biomarkers: predictive and prognostic, and disentangling them is a key challenge. Our first step is to phrase biomarker ranking in terms of optimizing an information theoretic quantity. This formalization of the problem will enable us to derive rankings of predictive/prognostic biomarkers, by estimating different, high dimensional, *conditional mutual information* terms. To estimate these terms, we suggest efficient low dimensional approximations. Finally, we introduce a new visualisation tool that captures the *prognostic* and the *predictive* strength of a set of biomarkers. We believe this representation will prove to be a powerful tool in biomarker discovery.

Keywords. Predictive biomarkers, prognostic biomarkers, mutual information

1. Introduction

We present an information theoretic approach to disentangle predictive and prognostic biomarkers. In clinical trials, a *prognostic biomarker* is a clinical or biological characteristic that provides information on the likely outcome irrespective of the treatment. On the other hand, a *predictive biomarker*, is a clinical or biological characteristic that provides information on the likely benefit from treatment. One of the key challenges in personalised medicine is to discover predictive biomarkers which will guide the analysis for tailored therapies, while discovering prognostic biomarkers is crucial for general patient care [9]. We should clarify that our work focuses on hypothesis generation (exploratory analysis), instead of hypothesis testing (confirmatory analysis) [5].

In our work we will focus on a clinical dataset $D = \{y_i, x_i, t_i\}^n_{i=1}$, where, y is a realization of a binary target variable Y, t is a realization of binary treatment indicator T (i.e. $T = 1$ if patient received experimental treatment, 0 otherwise), and x is a p-dimensional realization of the feature vector X, which describes the joint random variable of the p categorical features (or biomarkers). To make the distinction between prognostic and predictive biomarkers more formal we will follow a strategy introduced by various previous works [4, 5]. Let us assume that the true underlying model is the following logistic regression with up to second order interaction terms:

[1] Corresponding author, School of Computer Science, University of Manchester, Kilburn Building, Oxford Road Manchester, M13 9PL; E-mail: konstantinos.sechidis@manchester.ac.uk.

$$\text{logit}P(y=1|t,\boldsymbol{x}) = \alpha + \underbrace{\sum_{i=1}^{p}\beta_i x_i + \sum_{i,j=1}^{p}\beta_{i,j}x_i x_j}_{\text{Prognostic term}} + \gamma t + \underbrace{(\sum_{i=1}^{p}\delta_i x_i + \sum_{i,j=1}^{p}\delta_{i,j}x_i x_j)t}_{\text{Predictive term}}.$$

Covariates with non-zero β coefficients are prognostic, while with non-zero δ coefficients are predictive. Our work proposes an information theoretic framework for deriving two different rankings of the biomarkers, one that captures their *prognostic* strength, and one that captures their *predictive* strength. On top of that, we introduce a visualisation tool that captures both the *prognosticness* and the *predictiveness* of a set of biomarkers. This tool enables us to identify potentially undiscovered biomarkers, worthy of further investigation.

2. Background on Biomarker Ranking

Here we connect the problem of biomarker discovery with the machine learning problem of feature selection and the clinical trials problem of subgroup identification.

2.1. Prognostic Biomarker Discovery and Feature Selection

We now demonstrate that the problem of selecting *prognostic biomarkers* is equivalent to feature selection using a supervised dataset $\{y_i,x_i\}^n_{i=1}$. There are many different methods for feature selection, but we will focus on information theoretic approaches, where, firstly we *rank* the features and then we *select* the top-k ones that contain most of the useful information. The underlying objective function is to find the smallest feature set X^* that maximizes $I(X^*;Y)$, or in other words that the shared information between X^* and Y is maximized. Brown et al. [2] derived a greedy optimization process which assesses features based on a simple scoring criterion on the utility of including a feature. At each step we select the feature X_k that maximizes the conditional mutual information (CMI): $J^{\text{CMI}}(X_k) = I(X_k;Y|\mathbf{X}_\theta)$, where $\mathbf{X}_\theta$ is the set of the features already selected. As the number of selected features grows, the dimension of $\mathbf{X}_\theta$ also grows, and this makes our estimates less reliable. To overcome this problem *low order* criteria have been derived. For example, by ranking the features independently on their mutual information with the class, we derive a ranking that takes into account the *relevancy* with the class label. Choosing the features according to this ranking corresponds to the *Mutual Information Maximization* (MIM) criterion; where the score of each feature X_k is given by: $J^{\text{MIM}}(X_k) = I(X_k;Y)$. This approach does not consider the *redundancy* between the features. By using more advanced techniques, we can take into account both relevancy and redundancy between the features themselves, *without* having to compute very high dimensional distributions. Brown et al. [2] showed that a criterion that controls relevancy, redundancy, conditional redundancy and provides a very good tradeoff in terms of accuracy, stability and flexibility is the *Joint Mutual Information* (JMI) criterion [11]: $J^{\text{JMI}}(X_k) = \sum_{X_j \in \mathbf{X}_\theta} I\,(X_k; Y|X_j)$.

Our aim is to explore how the above framework can been extended to be useful in clinical trial scenarios, i.e. dataset $D = \{y_i,x_i,t_i\}^n_{i=1}$. The extra treatment variable T provides interesting dynamics, but before showing our suggested extension, we will briefly present the literature on predictive biomarkers and subgroup identification.

2.2. Predictive Biomarker Discovery and Subgroup Identification

The problem of deriving *predictive biomarkers* is closely related to the problem of subgroup identification [5]. In clinical trials, patient populations cannot be considered homogeneous, and thus the effect of treatment will vary across different subgroups of the population. Exploring the heterogeneity of subject responses to treatment is very critical for drug development, which is underlined by a draft FDA (Food and Drug Administration) guidance [9]. As a result, consideration of patient subgroups is necessary in multiple stages of trial development. Berry [1] gives the following definition: subgrouping is a partition of the set of all patients into disjoint subsets or subgroups and it is usually determined by a small number of measurable covariates, which are the predictive biomarkers. In the traditional subgroup identification problem, the set of predictive biomarkers is relatively small, i.e. 2-3 biomarkers [6].

In the literature, there are many different methods for subgroup identification. A popular one is *recursive partitioning* of the covariate space, using criteria that capture the interaction between T and Y [6, 7, 10]. Another solution builds upon the *counterfactual modelling* idea: firstly, by deriving a new variable for each patient that captures the treatment effect and then using this variable to select or rank the covariates. For example, Foster et al. [4] can be seen as exploring the covariate space which maximizes the odds-ratio between T and Y. In the following section, we will show that starting from a natural objective function, we can derive predictive biomarkers by exploring areas that maximize the mutual information between T and Y.

3. An Information Theoretic View on Biomarker Ranking

Our work extends the feature ranking framework from supervised to clinical trial data. The treatment variable T provides extra useful information, and a natural way to capture this is by the following criterion: to maximize the shared mutual information between the target Y and the joint random variable of the treatment T and the optimal feature set X^*, or in information theoretic notation: $X^* = \text{argmax}\ I(X_\theta T;Y)$. By using the chain rule [3], this objective can be decomposed as follows in the following way:

$$\boldsymbol{X}^* = \text{argmax}_{\boldsymbol{X}_\theta \in \boldsymbol{X}} I(\boldsymbol{X}_\theta T;Y) = \text{argmax}_{\boldsymbol{X}_\theta \in \boldsymbol{X}} \big(\underbrace{I(\boldsymbol{X}_\theta;Y)}_{\text{Prognostic term}} + \underbrace{I(T;Y|\boldsymbol{X}_\theta)}_{\text{Predictive term}} \big).$$

The first term, captures the features with prognostic power, while the second captures the features with predictive power. By optimizing these two terms independently we can derive two greedy optimization process, where at each step we select the feature X_k that maximizes the following terms: $J_{\text{Prog}}(X_k)=I(X_k;Y|\boldsymbol{X}_{\text{Prog}})$ and $J_{\text{Pred}}(X_k)=I(T;Y|\, X_k\boldsymbol{X}_{\text{Pred}})$, where $\boldsymbol{X}_{\text{Prog}}$ are the features already been ranked as prognostic, while $\boldsymbol{X}_{\text{Pred}}$ as predictive. As the number of selected features grows, the dimension of $\boldsymbol{X}_{\text{Prog}}$ and $\boldsymbol{X}_{\text{Pred}}$ also grows, and this makes the estimates less reliable. To overcome this issue, with the following theorem we derive low-order approximations, such as the one presented in Section 2.1.

Theorem 1. *The first two order approximations are given by:*

$$J^{1st}_{\text{Prog}}(X_k) = I(X_k;Y), \qquad J^{1st}_{\text{Pred}}(X_k) = I(T;Y|X_k).$$

$$J^{2nd}_{\text{Prog}}(X_k) = \sum_{X_j \in \boldsymbol{X}_{\text{Prog}}} I(X_k;Y|X_j), \quad J^{2nd}_{\text{Pred}}(X_k) = \sum_{X_j \in \boldsymbol{X}_{\text{Pred}}} I(T;Y|X_k X_j).$$

Proof sketches: For prognostic, the proof is identical to [2], while for the predictive we can prove these approximations by combining the results of [2] with the chain rule [3].

For example, by making assumptions similar to the ones of MIM, we can derive the 1^{st}-order criteria for deriving prognostic and predictive rankings respectively. These criteria do not consider interactions between features, and thus fail to capture the *redundancy*. To overcome this limitation so we can use higher order criteria, such as JMI, which explores 2^{nd}-order interaction terms between features.

4. Predictive–Prognostic (PP) Graphs

We now present a visualisation tool that captures both the *prognostic* and the *predictive* power of a set of biomarkers (PP-graphs). We believe that this representation will provide useful information over both the prognostic and predictive power of each biomarker, and it will be helpful for controlling false discoveries in clinical trials. For example, in subgroup-identification (Section 2.2), we define interesting subgroupings by using predictive biomarkers. Many methods, such as the counterfactual modelling, i.e. Virtual twins suggested by [4], derive as predictive, biomarkers that are strongly prognostic. Using a PP-graph we get more insight over the prognostic and predictive power of each biomarker and this may help in eliminating this type of errors.

Now we will show these graphs through a motivating example. We will use the same data generation model as in [4]. Let us assume that we simulate randomized trials with 1000 patients, and the Xs are generated as independent $X_j \sim N(0,1)$, $j = 1 \ldots 15$. We consider logit models for data generation:

$$\text{logit}P(y = 1|t,x) = -1 + 0.5(x_1 + x_2 - x_7 + x_2\, x_7) + 0.1t + 1.5tI(x_1 > 0 \cap x_2 < 0 \cap x_3 > 0).$$

The patients with $(x_1 > 0 \cap x_2 < 0 \cap x_3 > 0)$ will have an enhanced treatment effect. As a result the three variables, X_1, X_2 and X_3, are the predictive biomarkers. Furthermore, X_1,X_2 and X_7 are the three prognostic biomarkers and the other nine biomarkers are irrelevant.

Figure 1 shows three PP-graphs. In the x-axis, we have the normalised score of each biomarker derived by a prognostic ranking. We normalised scores to take values from $[0,1]$, where 1 is the score for the most-prognostic biomarker. In the y-axis, we have the normalised scores for the predictive ranking. The red area (vertical shaded region) represents the top-k prognostic-biomarkers, while the green (horizontal shaded region) the top-k predictive, for these specific PP-graphs we used $k = 3$, which corresponds to the score cut-off value of $(p - k)/p = (15 - 3)/15 = 0.80$. The intersection of these two areas – orange area (top right shaded corner)– should contain the biomarkers that are both prognostic and predictive. We plot the average predictive/prognostic rankings over 100 sample datasets, using Virtual-twins [4] and our two approaches suggested in Theorem 1. For estimating mutual information, the features were discretized in 4 equal width bins. As we observe, Virtual-twins, tends to push a prognostic biomarker (X_7) into the predictive area –*false positive*. The 1^{st}-order approach classifies X_1 only as prognostic and not as predictive –*false negative*. While, our 2^{nd}-order criterion distinguishes biomarkers perfectly.

5. Conclusions and Future Work

In this work, we focused on disentangling rankings of the biomarkers that quantify their predictive and their prognostic power. We presented an information-theoretic approach, where we started from a clearly specified objective function and we suggested lower

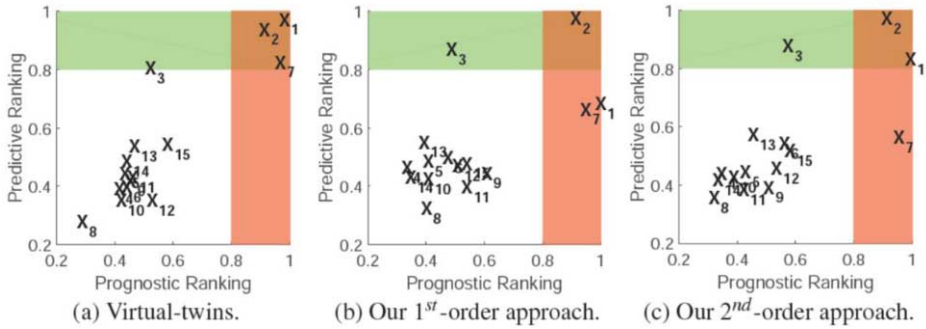


Figure 1. P-P graphs when: X_1,X_2 and X_3 are truly predictive, X_1,X_2 and X_7 are truly prognostic, and the rest nine biomarkers are irrelevant. Note that our 2^{nd}-order approximation distinguishes perfectly between predictive and prognostic.

order approximations. Lastly, we introduced a new graphical representation that captures the dynamics of biomarker ranking.

While in this paper we present results only from simulated datasets, we have a forthcoming work that applies our methodologies to a dataset coming from a real clinical trial on the progression free survival in a lung cancer study [8]. Our preliminary results confirm that the presence in the tumor of a mutation of the epidermal growth factor receptor (EGFR) gene is a predictive biomarker for two different treatments: gefitinib versus carboplatin–paclitaxel.

Another interesting future direction is to improve the interpretability of the P-P graphs. For example, in the 1^{st}-order approach, instead of plotting the ranking score of each biomarker, we can plot a p-value, derived from a univariate testing of whether the biomarker is predictive or prognostic.

References

[1] D. A. Berry. Subgroup analyses (letter). *Biometrics*, 46(4):1227–1230, 1990.
[2] G. Brown et al. Conditional Likelihood Maximisation: A Unifying Framework for Information Theoretic Feature Selection. *JMLR*, 13:27–66, 2012.
[3] T. M. Cover and J. A. Thomas. *Elements of information theory*. Wiley, 2006.
[4] J. C. Foster et al. Subgroup identification from randomized clinical trial data. *Statistics in Medicine*, 30(24):2867–2880, 2011.
[5] I. Lipkovich and A. Dmitrienko. Strategies for identifying predictive biomarkers and subgroups with enhanced treatment effect in clinical trials using SIDES. *Journal of Biopharmaceutical Statistics*, 24(1):130–153, 2014.
[6] I. Lipkovich et al. Subgroup identification based on differential effect search. *Statistics in Medicine*, 30(21):2601–2621, 2011.
[7] W.-Y. Loh et al. A regression tree approach to identifying subgroups with differential treatment effects. *Statistics in medicine*, 34(11):1818–1833, 2015.
[8] T. S. Mok et al. Gefitinib or Carboplatin–Paclitaxel in Pulmonary Adenocarcinoma. New England Journal of Medicine, 361(10): 947-957, 2009.
[9] Stephen J. Ruberg and Lei Shen. Personalized medicine: Four perspectives of tailored medicine. *Statistics in Biopharmaceutical Research*, 7(3):214–229, 2015.
[10] X. Su et al. Subgroup analysis via recursive partitioning. *JMLR*, 10:141–158, 2009.
[11] H. H. Yang and J. Moody. Data visualization and feature selection: New algorithms for nongaussian data. In *NIPS*. 1999

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-146

IntegrIT - Towards Utilizing the Swedish National Health Information Exchange Platform for Clinical Research

Maria HÄGGLUND[a,1], Therese SCOTT DUNCAN[a], Karin KAI-LARSEN[b], Gunilla HEDLIN[b] and Ingvar KRAKAU[c]
[a] *Health Informatics Centre, Karolinska Institutet*
[b] *Institute of Environmental Medicine, Karolinska Institutet*
[c] *Department of Medicine, Karolinska Institutet*

Abstract. This paper describes how the Swedish national Health Information Exchange platform can be used to facilitate clinical research in the future. Different e-services for different user groups are being developed using a user-centered design approach. The main user groups are study participants, clinical researchers and healthcare professionals. The different e-services are based on an in-depth analysis of the clinical research process, and the main identified needs relate to recruitment of study participants, access to clinical data from different sources as well as improved tools for patients' self-reporting. The national Swedish HIE platform has the potential to enable a seamless connection between patients/citizens as study participants, health care professionals and everyday clinical work and clinical researchers in both academia and industry.

Keywords. Clinical Research Informatics, eHealth, User-Centered Design

1. Introduction

Clinical research is an important part of healthcare and clinical research informatics provides essential tools to support and reshape the landscape of clinical research [1][2]. Clinical research informatics is a relatively young field of research, and there are many challenges that need to be addressed, whereof data access and recruitment issues are often mentioned as some of the most important ones [3][4]. Many international studies are focusing on addressing the challenges of gaining access to clinical data, from e.g. the electronic health records (EHR), for research purposes [5][6].

In parallel with the progress within clinical research informatics, participatory health is a growing area in which individuals are using health social networks, smartphone health applications, and personal health records to achieve positive health outcomes. In Sweden, a national eHealth infrastructure [7] is currently being implemented enabling health data to be accessible across care providers and to patients.

Mobile technology and self-tracking devices are also radically impacting the way we collect, use and share health related data, and when patients pool their data together powerful data sets are created that can be of great importance to clinical research. In the

[1] Corresponding author, Karolinska Institutet, Tomtebodavägen 18A, 17177 Stockholm, Sweden; E-mail: maria.hagglund@ki.se

US, the Precision Medicine Initiative is aiming to explore how such large data sets capturing can help create treatments that are adapted to individual variability in genes, environment, and lifestyle for each person [8]. In Sweden, several initiatives are exploring the opportunities of utilizing the national eHealth infrastructure for secondary purposes, e.g. to support clinical research. In this paper, we aim to describe how the Swedish national Health Information Exchange (HIE) platform can be used to facilitate clinical research in the future.

2. Methods and Materials

IntegrIT is a 2-year project with the goal to develop new services and improve routines for patient recruitment to clinical research projects, systematic, ongoing collection and analysis of structured clinical data for research directly from medical records, and integration of clinical data with patients' self-reporting and data from biobanks and registries. To be able to access data from many different EHR systems currently used within Sweden, the solutions are based on the national HIE platform and will be part of a new national research infrastructure. The infrastructure will be integrated with the already available national virtual patient portal 1177 Vårdguiden (http://www.1177.se/) and the corresponding portal for healthcare professionals. The purpose is to promote clinical research in daily care while reducing administrative time for both researchers and study participants.

2.1. The Swedish National Health Information Exchange Platform

Sweden has chosen to implement a national HIE platform to facilitate the communication between different health information systems and eHealth services. The national HIE platform enables a single point of connectivity for client applications, making all Swedish EHRs appear as a national, virtual EHR. Client applications may be targeted for patients, professionals, researchers, payers, byers and follow-up. The national HIE platform allows exchange of health care data between different health information systems (HIS) according to nationally defined *service contracts*. Rather than having direct integration between HIS, all integration is with the national HIE platform which then redirects requests for information and transactions to the appropriate system. In summary, the national HIE platform forwards the request message from a system or a service to the appropriate source systems, often e.g. EHR systems used by different care providers and returns the response, which may be aggregated from multiple sources [9].

2.2. User-Centered Design

A user-centered design approach [10] has been applied in the IntegrIT-project to ensure that the solutions are usable and adapted to the context of the different user groups. The main user groups were identified as study participant, clinical researchers and health care professionals, and an in-depth analysis of their needs was performed using qualitative methods; interviews [11] and contextual inquiries[12]. Based on the contextual inquiries with clinical researchers, a hierarchical task analysis [13] was performed which formed the basis for the proposed functionality for different user groups.

3. Results

In this paper we provide an overview of the e-services developed in the IntegrIT project. As an important part of the context description, we first present the different user groups, and then continue to describe the services proposed in the project.

The stakeholders are many in clinical research, and there is great interest in a clinical research infrastructure. Within the IntegrIT project we have chosen to focus on 3 main user groups which are described further in table 1.

Table 1. Description of IntegrIT's main user groups.

User group	Description
Clinical researchers	Clinical researchers (physicians, nurses, physiotherapists etc). Work situation and incentives differ a lot between researchers in the hospital and primary care contexts, as well as clinical researchers in industry. Research nurses is a specific sub-group of clinical researchers who work almost exclusively with clinical research but rarely as Principal Investigator (PI)
Study participants	Patients are the main type of study participants in clinical research, but we also include next-of-kin in this category (e.g. parents of children participating in clinical studies). In addition to patients, healthy volunteers are also an important sub-group that could have an interest in IntegrIT's e-services.
Health care professionals	In many clinical studies that are performed in daily care, health care professionals are required to contribute to the study by e.g. recruiting study participants and collecting study data. In this category we also include healthcare managers who need to approve the participation of staff as well as access to clinical data.

As part of the user needs analysis, a hierarchical task analysis was performed based on the contextual inquiries made with clinical researchers. The activities or tasks researchers go through within a clinical study was identified and broken down in several layers. Here, we only present the top layer which corresponds to the overall phases in a clinical study, figure 1. Whereas there were many tasks identified in all phases, the analysis indicated that IntegrIT's e-services could be most useful in phase 3 and 4 – preparing and conducting the study.

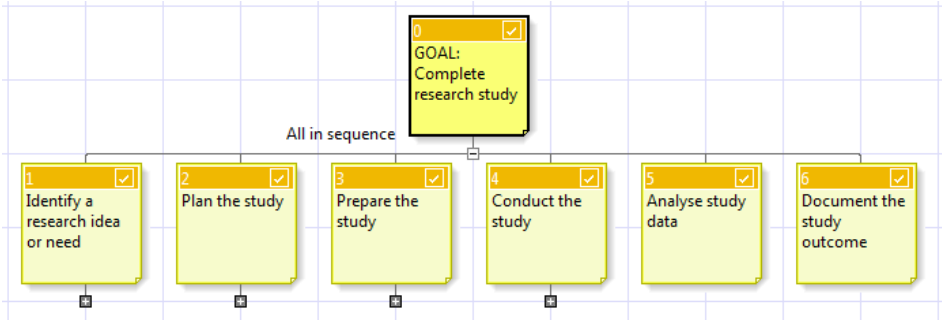


Figure 1. Top layer of the hierarchical task analysis

The analysis of activities in the clinical research process formed the basis for identifying the core functionality needed to create useful tools to support different users through the clinical research process (figure 2).

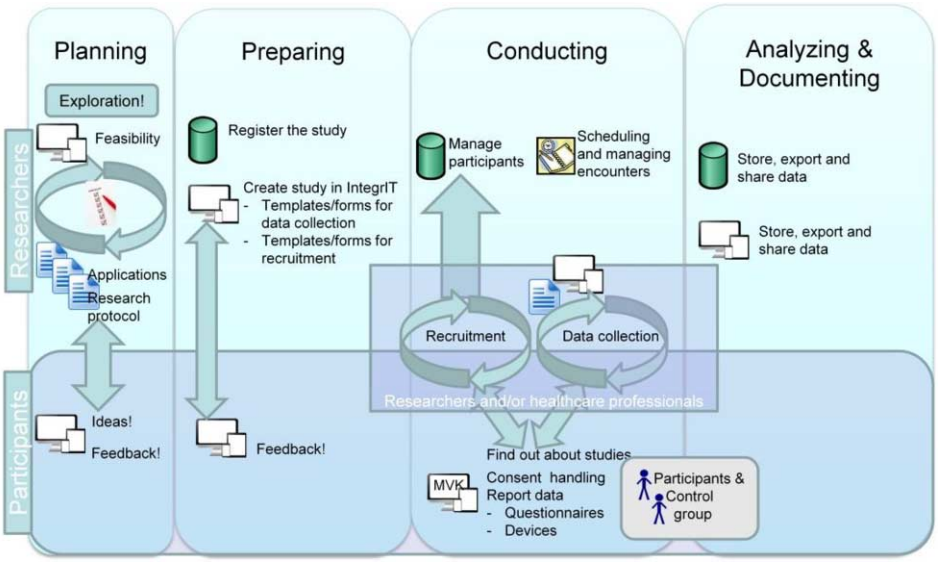


Figure 2. An overview of the services proposed by IntegrIT

In the first stage of the project, an initial set of e-services was developed based on the main needs identified as well as the feasibility of providing the functionality through the national HIE platform.

Table 2. Description of IntegrIT's main e-services.

e-service	Description
Study Guide	A study guide was developed based on the need to support less experienced researchers in planning and preparing a study. The study guide e.g. describes what approvals are needed, as well as where and how to obtain them.
Study service – Researchers	Clinical researchers can log in to register their studies. Current functionality includes support for recruiting participants through the national patient portal and to set-up recruitment rules for matching with EHR data, to handle study participants from first contact to inclusion, to handle informed consent (electronically or manually) and to collect data from EHRs (with patient and care provider consent). Researchers can also invite other investigators and add further study sites.
Study service - Study participants	The study service for citizens/patients is intended to be reached through the national patient portal with secure log-in. After log-in, the user can search for available research studies in a study catalogue, use an EHR matching service to find studies that could be relevant based on the information in the EHR, contact studies that are interesting with a request to participate, receive information and provide informed consent (and have access to consents given and study information in retrospect).
Study service – Care provider	This service contains two main functions; (1) to approve that a study can gain access to EHR data, and (2) a matching function similar to the one used by the patients themselves but intended to be used by healthcare professionals together with the patient during an encounter. A so called Clinical Trial Alert [5] function is also under development in collaboration with EHR vendors.

In addition, an iPhone app for self-reporting of study data has been developed based on the Apple Research Kit [14], for study participants to be able to report their allergy symptoms on a daily basis during the allergy season. The app was used in a study during the summer 2016 and evaluation from a user perspective is ongoing. The app is not yet integrated with the study services described above.

4. Discussion and Conclusion

The national Swedish HIE platform has the potential to provide not only useful tools for clinicians (e.g. the national patient overview) and patients (e.g. online access to one's own EHR) but also to enable a more seamless connection between patients/citizens as study participants, health care professionals and everyday clinical work and clinical researchers in both academia and industry. In this paper, we have presented the initial steps of developing a national clinical research infrastructure based on the national HIE platform. The e-services developed in the first stage of the project have undergone formative evaluation through usability testing, however further pilot testing in clinical studies is required to ensure that both infrastructure and e-services are sufficient to support clinical research in the future.

Acknowledgments

The project "IntegrIT" is supported by VINNOVA – Swedish Agency for Innovation Systems (2014-00736).

References

[1] R. Richesson and J. Andrews, *Clinical Research Informatics*. London: Springer-Verlag, 2012.
[2] M. G. Kahn and C. Weng, "Clinical research informatics: a conceptual perspective.," *J. Am. Med. Inform. Assoc.*, vol. 19, no. e1, pp. e36-42, 2012.
[3] P. J. Embi and P. R. O. Payne, "Clinical Research Informatics: Challenges, Opportunities and Definition for an Emerging Domain," *J. Am. Med. Informatics Assoc.*, vol. 16, no. 3, pp. 316–327, 2009.
[4] L. Richter-Sundberg, M. E. Nystrom, I. Krakau, and C. Sandahl, "Improving treatment of depression in primary health care: a case study of obstacles to perform a clinical trial designed to implement practice guidelines," *Prim Heal. Care Res Dev*, vol. 16, no. 2, pp. 188–200, 2015.
[5] J.-F. Ethier, V. Curcin, A. Barton, M. M. McGilchrist, H. Bastiaens, A. Andreasson, J. Rossiter, L. Zhao, T. N. Arvanitis, A. Taweel, B. C. Delaney, and A. Burgun, "Clinical data integration model. Core interoperability ontology for research using primary care data.," *Methods Inf. Med.*, vol. 54, no. 1, pp. 16–23, 2015.
[6] F. D. R. Hobbs, "Envisioning a Learning Health Care System :," *Ann Fam Med*, vol. 10, pp. 54–59, 2012.
[7] N. Lundberg, S. Koch, M. Hägglund, P. Bolin, N. Davoody, J. Eltes, O. Jarlman, A. Perlich, V. Vimarlund, and C. Winsnes, "My Care Pathways - creating open innovation in healthcare," *Stud Heal. Technol Inf.*, vol. 192, pp. 687–91, 2013.
[8] "Precision Medicine Initiative." [Online]. Available: https://www.nih.gov/precision-medicine-initiative-cohort-program .
[9] "INERA - arkitektur och infrastruktur." [Online]. Available: http://www.inera.se/ARKITEKTUR--INFRASTRUKTUR/. [Accessed: 06-Nov-2016].
[10] ISO 9241-210, "Ergonomics of human-system interaction. Part 210: Human-centred design for interactive systems," 2010.
[11] N. K. Denzin and Y. S. Lincoln, "Handbook of Qualitative Research." Sage, Thousand Oaks, 2000.
[12] H. Beyer and K. Holtzblatt, *Contextual Design: defining customer-centered systems*. San Fransisco: Morgan Kaufmann Publishers, 1998.
[13] N. A. Stanton, "Hierarchical task analysis: Developments, applications, and extensions," *Appl. Ergon.*, vol. 37, no. 1, pp. 55–79, 2006.
[14] "Apple Research Kit." [Online]. Available: https://www.apple.com/researchkit/.

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-151

Introducing a Method for Transformation of Paper-Based Research Data into Concept-Based Representation with openEHR

Birgit SAALFELD[a], Erik TUTE[a], Klaus-Hendrik WOLF[a] and Michael MARSCHOLLEK[a]
[a]*Peter L. Reichertz Institute for Medical Informatics University of Braunschweig - Institute of Technology and Hannover Medical School , Germany (PLRI)*

Abstract. Combining research data and clinical routine data is a chance for medical research. We present our method for the transformation of paper-based research data into a concept-based representation. With this representation the study data from research projects can be combined with data from clinical tools with less integration effort. We applied and verified our method using data from a current research study. In this paper we also show our main challenges and lessons learned. Clinical assessment data and study diaries from a long term study (n=24, 3 months observation time each, 17 different clinical assessments) stored on paper were used as the data set. An openEHR-based electronical health record platform was adapted for acquisition and representation of the research data. To avoid transcription errors, the data was entered twice by different student assistants. A third compared and corrected both data sets. Content models (17 archetypes and five templates from openEHR concept) based on clinical assessments were created manually. Web forms for data entry were created automatically on the basis of this concept-based content models. Additionally, form functionalities to support data entry and comparison were implemented. In total, 829 compositions were entered by the student assistants. With our developed method, we are able to represent the study data in a clinical concept-based platform, which means less integration effort for access and processing of research and clinical data. Some minor difficulties occurred during the process. All in all, adapting routine tools, like the EHR platform, seems to be convenient to deal with research data.

Keywords. openEHR, data representation, concept-based, data integration

1. Introduction

Nowadays, it is more and more common to publish research data to increase the reproducibility of research results and give other researchers the chance to do further research on this data. In the research project GAL-NATARS [1] it is also intended to make the obtained data publicly available.

At Hannover Medical School in the project HaMSTR [2] the research focuses on a concept-based provisioning system for making clinical routine data for research purposes available. Concretely, a clinical data warehouse prototype with an openEHR

based data storage layer is deployed. Combining research data and clinical routine data is considered to have plenty of potential for medical research.

This led to the idea of testing a concept driven EHR-platform for the acquisition and representation of research data.

Being aware of the fact that nowadays probably most research data is captured digitally and integration of this data is of great importance, in this paper we present our method for the transformation of paper-based research data into concept-based representation on the example of the assessment results from the GAL-NATARS study. Furthermore, we want to share our experiences and lessons learned.

2. Material and Methods

2.1. Data

The Peter L. Reichertz Institute had a prolective, multicentre observational trial about multimodal activity monitoring for home rehabilitation, called GAL-NATARS [1]. Three clinics took part in the trial. All in all 24 subjects were observed[1]. All of them are aged over 70 and had a fracture affecting their mobility. After geriatric rehabilitation the single-living people were equipped with different sensor systems. During the three months lasting observation time a study nurse visited the subject once a week to check the technical system and do assessments.

At the begin and at the end of observation 16 assessments[2] were done in two days each (T0a, T0b, T3a, T3b). Every 4th week a subset of those assessments (T1, T2) was done to record the subjects recovery development. Furthermore every week the BMI, the fall risk, and their pain experience were documented (T+), see figure 1. Additional, every day the subjects themselves or their relatives made a short diary entry. In a structured way they wrote down the personal comfort of the subjects, potential occurred falls, visitors in their flat and in an additional field for notes the extraordinary events that happened that special day.

2.2. Platform

Among other things, openEHR provides standards for clinical content models and connection points to terminology - the openEHR archetypes and templates.

Every archetype is the representation of a medical concept in a semantically unambiguous and machine readable form, like the blood pressure or the body mass index. Archetypes are intended to be reused in different use cases. Templates describe the composition of information units like archetypes for a specific use case. [3] Beyond that it offers a concept-based query language named Archetype Query Language (AQL) and it defines interfaces, like a REST-API, to access and modify the data. openEHR is used in several countries for example Australia, Norway, Sweden and the

[1] In the paper [1] only 13 subjects are mentioned, because the study was still running when the paper had been published

[2] Barthel Index, Body Mass Index, Clock Drawing Test, Falls Efficacy Scale, Geriatric Depression Scale, Grip Strength Meausurement, Activities of Daily Live, Mini Mental State Examination, Mini Nutritional Assessment, Prevention of Falls Network Europe, Questionnaire on Technical Readiness, Short Physical Performance Battery, Social Situation, Timed Up and Go, Tinetti, Visual Analog Scale

UK [4]. The *Think!EHR Platform* by Marand implements the concept of a semantically enabled health computing platform on the basis of openEHR. The platform is used among others in Moscow City as an Integrated Medical Information System and in Slovenia as a basis for the National health information exchange network. [5]

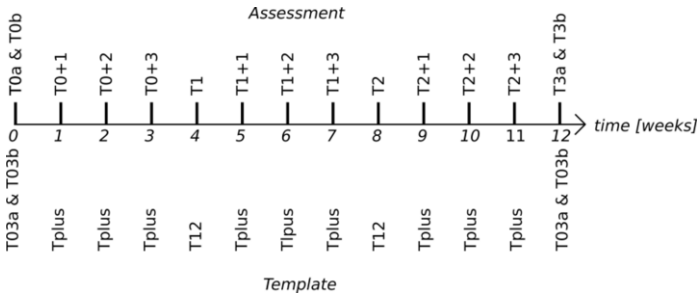


Figure 1. Measurement points and the corresponding templates

2.3. Methods

First the concepts behind the medical assessments are modelled as archetypes. In the next step the archetypes become combined in templates, like they were combined on the single assessment days at the subjects home.

The form builder from *Marand* is used to create a form description from a template definition automatically. This definition can be extended manually with tags to bind for example custom functionalities special form elements. Those functions can be implemented using JavaScript.

For manual data input a tool was created based on JavaScript and the REST-API offered by the *Think!EHR Platform*. The form description is used in the JavaScript renderer library from Marand for an automatic visual representation. The tagging mechanism was used with JavaScript to implement functionality like summing up scores or calculating the body mass index.

For validation all data is entered twice into the system. Both entry processes are independent from each other. Afterwards a comparison of both data sets has to be performed. Therefore a comparison tool to highlight the differences should be used. Since no tool exists to compare openEHR compositions, retrieved as JSON-objects from the REST-API, a self-developed solution has to be created. The self-developed tool compares JSON-export files from every subject and every assessment and colors the spotted differences to support the comparing person.

During the comparison process the data sets will become equalized, so that both data sets represents the original data and to avoid transmission errors from paper to openEHR.

3. Material and Methods

The *Archetype Editor* from the openEHR Foundation Version 2.8.972 Beta was used to build the 16 evaluation archetypes and one composition archetype. Since one central idea of detailed clinical models is using the same archetypes in different settings and reusing existing ones, we looked for existing openEHR archetypes and other detailed

clinical models [2] und [6]. We did not find models that allowed us to exactly represent our paper forms used in the GAL-NATARS project. To prevent distortion of the research data we decided to create our own models. For template creation the *Template Designer* 2.6.1214 Beta was used. All in all five different templates were created, see table Table 1. There is also shown, how many compositions were created by the student assistants for the different template types.

Table 1. Types of templates

Template	Use	Number compositions
T03a	First of two days at starting and ending observation time	44
T03b	Second of two days at starting and ending observation time	44
T12	Big assessment after one and two months	47
Tplus	Weekly, in case there is no bigger Template to be done	177
TDiary	Daily	23 or much more

For digitalization of the data two student assistants retyped the data independent from each other into the openEHR-platform. They used the Google Chrome Browser (Version 54.0.2840.71 m) and entered the data for 24 subjects, each with up to 15 assessment days. A feature of archetypes are constraints. They may be useful to ensure that the necessary data is filled in. In our first archetypes all fields had the occurrence constraint to be exactly one. This turned out to be a problem, when the paper assessments were not filled completely. So we had to adapt all archetypes to work without that preferred constraint.

After entering the data (829 compositions in total) a comparison of both data sets was done by a third student assistants. The comparison for the assessment based archetypes worked without difficulties, but the diary entries were special.

From each student assistants the diary entries (each day as instance of an archetype) were entered in different orders, but with correct dates. Thus the comparison tool spotted differences between both identical data sets due to the different item order. An additional sorting function had to be used before the comparison.

For the export of data we used AQL. When a query is executed with archetype fields this can look very confusing, because attributes sometimes cannot be called by the name, but by the attribute identifier, like *att0035*.

4. Discussion

With our developed method, we are able to represent the study data in a clinical concept-based platform, which means less integration effort for access and processing of research and clinical data. However, there were some challenges and lessons learned, which are presented in the following paragraphs.

The most complex task was building the content models (archetypes and templates). The student assistants entered different numbers of diary entries in one composition. So, one student assistants needed more compositions than the other. That is why, in total, an odd number of compositions exists. Constraints for single fields are useful during data entry to enhance data quality. Using constraints in the process of digitalization can cause problems, except when there is a possibility to explicitly define a value as missing.

Combining data from research and routine is possible in two different ways. The first possibility is dealing with both data sets separately using tools from research or routine respectively, but this results in a bigger effort integrating them. The other option is adapting the tools from one domain to deal with requirements from both and have less effort integrating them. Our experiences showed no major problems for adaptation of a tool from routine to the research requirements. Since we did not benchmark our openEHR-based tool chain against professional tools for clinical studies that may support form generation, double data entry and comparison, we are not able to state that one approach is superior. Finally, experiencing no major problems in one example implementation cannot prove the absence of potential problems within that approach in general.

5. Conclusion

We developed a method for the transformation of paper-based research data into concept-based representation using the example of the assessment results of the GAL-NATARS study. The study data from real subjects is now digitalized for further research. Therefore, the entered data will be made publically available, along with the related archetypes and templates. Some minor difficulties occurred during the process, but all in all adapting the routine tools to deal with research data seems to be convenient.

6. Acknowledgement

We would like to acknowledge the Lower Saxony research network “Design of Environments for Ageing” supported of the Lower Saxony Ministry of Science and Culture through the “Niedersaechsisches Vorab” grant programme (grant ZN 2701) for opportunity for use the data. We also want to thank the student assistants Alexander Harms, Aylin Uzun and Christina Valtin for the support with the manual data input.

7. References

[1] Marschollek M, Becker M, Bauer JM, Bente P, Dasenbrock L, Elbers K, et al. Multimodal activity monitoring for home rehabilitation of geriatric fracture patients--feasibility and acceptance of sensor systems in the GAL-NATARS study. Inform Health Soc Care. 2014;39:262–71. doi:10.3109/17538157.2014.931852.

[2] Haarbrandt B. Hannover Medical School Translational Research Framework (HaMSTR). https://plri.de/en/forschung/projekte/hannover-medical-school-translational-research-framework-hamstr. Accessed 04-November-2016.

[3] K. Atalag, T. Beale, R. Chen, T. Gornik, S. Heard, and I. McNicoll. openehr. http://www.openehr.org/resources/white_paper_docs/openEHR_vendor_independent_platform.pdf. Accessed 01-November-2016.

[4] openEHR Foundation. Deployed solutions. http://www.openehr.org/who_is_using_openehr/healthcare_providers_and_authorities. Accessed 04-November-2016.

[5] Marand d.o.o. Marand Homepage. http://www.marand.com/. Accessed 04-November-2016.

[6] CIMI. Website der Clinical Information Modeling Initiative. http://opencimi.org/about_cimi. Accessed 04-November-2016.

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-156

The 'PEARL' Data Warehouse: Initial Challenges Faced with Semantic and Syntactic Interoperability

Samhar MAHMOUD[a,1], Andy BOYD[b], Vasa CURCIN[a], Richard BACHE[a], Asad ALI[b], Simon MILES[a], Adel TAWEEL[a], Brendan DELANEY[c] and John MACLEOD[b]
[a] *King's College London*
[b] *University of Bristol*
[c] *Imperial College London*

Abstract. Data about patients are available from diverse sources, including those routinely collected as individuals interact with service providers, and those provided directly by individuals through surveys. Linking these data can lead to a more complete picture about the individual, to inform either care decision making or research investigations. However, post-linkage, differences in data recording systems and formats present barriers to achieving these aims. This paper describes an approach to combine linked GP records with study observations, and reports initial challenges related to semantic and syntactic interoperability issues.

Keywords. Data Linkage, Electronic Health Records, PEARL, ALSPAC

1. Introduction

Cohort studies increasingly implement comprehensive record linkage programs to retrospectively and prospectively collect information on study participants. To effectively combine linked data with survey data it is necessary to bring these disparate data into one single, heterogeneous representation. The Project to Enhance ALSPAC through Record Linkage (PEARL) is developing a data processing and warehousing solution (DWH) to help resolve these issues. The approach typically taken is to create a single DWH constructed according to a well-defined data model. In a clinical context, electronic patient records (EPRs) are typically arranged as an event sequence. In contrast, observational studies tend to collate observations in wide, flat, file structures; where each record represents a participant, and each file a data collection exercise. The PEARL DWH combines both clinical and self-reported information into a single event record. This format is familiar to clinical researchers, is novel in a cohort context, and contrasts with other contemporary approaches [1,2]. This format allows users to efficiently extract data using standard querying languages available in routine analytical software. This paper describes: the chosen data model; data pipeline workflows that combine EPRs and self-reported participants' data of the Avon Longitudinal Study of Parents and Children (ALSPAC); and, the methods developed to overcome interoperability challenges.

[1] Corresponding author, Division of Health & Social Care Research, Faculty of Life Sciences & Medicine, King's College London, The Strand, London, WC2R 2LS; E-mail: samhar.mahmoud@kcl.ac.uk.

2. Background

ALSPAC is a longitudinal birth cohort study collecting information of participants' life course exposures, and health, social and well-being outcomes. ALSPAC recruited pregnant women - living in, and around, the City of Bristol - due to deliver between 01/04/91 and 31/12/92 [3]. An initial total of 14,062 live-born children were enrolled. Data is collected via questionnaires, study assessment visits, biological and 'omic characterisations (see: www.bristol.ac.uk/alspac/researchers/access/). PEARL (PI John Macleod) was designed to complement these self-reported data through the secondary use of linked routinely collected records. Ethical approval was obtained from the ALSPAC Ethics and Law Committee and NHS Research Ethics Committee (Ref: Haydock 10/H1010/70). To help ensure acceptable data usage, PEARL implemented a 'Data Safe Haven' governance framework [4]. The safe haven incorporates a 'UK Secure eResearch Facility' (UKSeRP) developed by the Welsh Farr Institute as a secure data repository and analysis platform. In 2013 Boyd and Macleod, with Egton Medical Information Systems (EMIS) Ltd and Apollo Medical Systems (Apollo) Ltd, used this framework to extract a pilot 3,166 EPR instances, relating to 2,249 ALSPAC participants, from 181 General Practices. An exemplar use-case was identified to test the functionality of the data warehouse. The use case – an investigation into genetic and environmental influences on asthma – will use linked EPRs to assess: i) clinical validation of self-reported data; ii) impact of prescribed treatment; and, iii) value of EPR data in missing data methods. The initial data migrated into the data warehouse relate to this use case.

3. Method

Within clinical data, we can distinguish between existential facts and value-bearing facts. Existential facts record only the fact that something occurred e.g. a diagnosis. Value-bearing facts record not just an occurrence but also some value associated with it such as a BMI reading and the corresponding value and unit of measurement. Three types of structured value have been identified in both clinical and non-clinical data and we use the ISO21090 data types [5], which are: Physical quantity (ISO21090 type PQ) where there is a numeric (scalar) quantity and a unit of measurement; Coded Ordinal (ISO21090 type CO) where there is an ordinal scale of values to which both a numeric value and a meaning (a code) are assigned; and Coded Value (ISO21090 type CD) where there is no numeric value or order over a range of values.

An event-based model is chosen, since we believe that all data sources can be rendered into a sequence of events. Thus, each data point is recorded as a distinct fact, with a single timestamp relates to precisely one individual. GP data sources already use an event-based representation. Equally, a self-reported questionnaire can be viewed as a series of events where each answer to a question is a single event. Some existing, and widely used, standards define conventions for the exchange of healthcare data between clinical sites. Four of these adopt an event-based representation: HL7 Reference Information Model (RIM, www.hl7.org); OpenEHR (www.openehr.org); Continuity of Care Record (CCR); and Continuity of Care Document (CCD). CCD and CCR have significant limitations for longitudinal studies, in that they do not support non-numeric values (CO and CD) - such as those generated by most questionnaire questions - and that for PQ data elements only a single reading is actually stored. Both OpenEHR and HL7 RIM support all the above data types. However, the availability of a database schema

with an inbuilt provenance model made the HL7 option the obvious one. We have also adopted the ISO 21090 data types known as the Constrained Information Model (CIM) [5], which has not, to date, been used to store non-clinical data.

3.1. The HL7 RIM-based Data Model

The main classes in the data model are, as shown in Figure 1: Subject – contains details of birth, death and administrative gender; Clinical Statement – records events such as: Observations – diagnoses, reporting of symptoms, measurements and any other observation, and Procedures – anything performed on the subject, of which a special case is Substance administration – medications, vaccines etc; and Organisation.

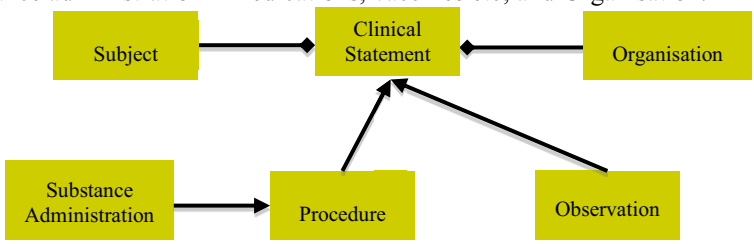


Figure 1. High-level class diagram of data model

For observations, there are 'code' and 'value' fields. For existential observations, the 'code' field is set to a coded value indicating a 'diagnosis' or 'finding'. The 'value' field is used to record the nature of the diagnosis. For value-bearing observations, the 'code' is again set to a coded value (e.g. indicating a laboratory test or vital sign). The 'value' field is used to record the reading. Here, the value may be of data type PQ, CO and CD. For data type CO, both a code and an (ordinal) numeric value are stored. For data type PQ a (scalar) value and unit of measurement are recorded. HL7 requires that all units of measurement must be recorded in UCUM (Universal Convention for Units of Measurement, www.dmd.nhs.uk). HL7 also requires that where a coding system is used it should be specified by means of its OID (Object Identifier). Therefore, each concept is specified by a code and OID defining the coding system and optionally a display name describing the concept and a free text string describing the coding system.

Table 1. Summary of GP Data from Two Sources

Attribute	EMIS Data	Apollo Data
No. Patient Record Instances	2787	379
No. Clinical Facts	400975	60668
No. Clinical Facts Processed	377681	53476
Non-medication Coding system	Read v2, SNOMED-UK	Read v2
Medication Coding System	SNOMED-UK, Local codes	Read v2
Units of Measurement	Mostly present	Missing
Format	OpenHR XML	CSV files

A majority of data (EMIS) is exported as XML files according to OpenEHR schema. A minority of the data (Apollo) is held in tab-delimited files. Table 1 summarises key features of the two data sources. Units of measurement, which are necessary for PQ-value facts, have two problems. First, it is missing in Apollo data and for some of EMIS data. Second, in EMIS data, the units are not specified in UCUM. Thus two further mappings are required: Non-standard unit to UCUM unit; and Read v2 codes for vital sign, lab test etc. and actual value to the UCUM unit. These were constructed on an ad hoc basis by PEARL staff and confirmed by clinical review.

3.2. Incorporating Questionnaire Data into HL7 RIM

Non-clinical data may relate to medical matters, such as self-reported (or parent/ carer) disease, symptom or medication use. It may also relate to non-medical matters such as housing, diet, lifestyle or opinions (e.g. do you like your teacher?). Of the non-clinical data available to PEARL, only questionnaire data has been formally considered to date. The Asthma use-case comprised of a subset of 217 questions from 34 ALSPAC mother and child questionnaires. All questions from the use case required a constrained response, where participants had to complete either a multiple-choice answer, with a prescribed set of answers or a numerical or timestamp reply.

Non-clinical data are recorded as observations, which do not differ substantially from clinicians recording survey responses, or consultation questions without any independent corroboration. Therefore, non-clinical questionnaire data with structured replies can be recorded as a value-bearing observation. Analysis of the asthma use case shows that all replies can be recorded as data types PQ, CD and CO.

The terms describing clinical facts are comprehensively recorded within structured coding systems (e.g. Read or SNOMED codes). However, survey responses do not have comprehensive code frames (although domain specific vocabularies' such as HASSET and MESH exist). Therefore, we have developed KASPER (King's Auxiliary System for Provisionally Encoding Records). PEARL uses SNOMED codes, where they exist, to code attributes such as height or body mass. Where questionnaire concepts are not supported by any terminology, a new KASPER code is created for the *root* concept. For data of types CO and CD, KASPER leaf codes are created for each of the possible replies that can be given. KASPER has its own OID registered with HL7. Figure 2 gives an example of how replies to questions are mapped to KASPER codes. The PEARL data is supplied with each subject being identified by the UKSeRP anonymised linkage field (ALF). This is unique to each subject and is the means by which records from different sources can be linked within the UKSeRP environment.

Code	Display Name	Value (code)	Value (number)	Display Name
S1003R	Disinfectant exposure in utero	S1002L3	3	Once per week
S1004R	Bleach exposure in utero	S1002L2	2	Most days
S1005R	Window cleaner exposure in utero	S1002L4	4	< Once per week

Figure 2. Mapping of Questionnaire Replies to KASPER codes

Questionnaires results are held in a CSV files. Each row represents one subject. The meaning of each variable is defined in the ALSPAC documentation. The data model requires that each attribute addressed either by a questionnaire question or composite score be represented by a code. Each variable in the asthma use case was checked against the ALSPAC documentation. Where the attribute had an existing SNOMED code e.g. body height, this was used. Otherwise a KASPER code was created. For CD and CO data types, it is necessary to create also leaf codes for each possible value. For variables of type PQ a unit of measurement in UCUM was specified.

3.3. Data Pipelines

The EPR extract and loading process has three stages. 1) Clinical facts are extracted and parsed into objects of type 'Fact'. 2) Data are cleaned and transformed to create a 'PEARL Fact' object, where: a 'Fact' with numerical values is set to a PQ-value observation; a 'Fact' with a code listed in the CD or CO mappings is set to a CD-value

or CO-value observation respectively; a 'Fact' with a code listed in the vaccine mappings is assumed to be a substance administration. 3) 'PEARL Facts' are loaded into the DWH.

4. Conclusion

There is a great importance for the linkage of clinical data to make it available for medical research. This paper describes an approach utilising an event based data model to combine participants' clinical data with their self-reported questionnaire data. A number of broad approaches exist, but mainly focus on clinical data of the same nature [7], or choose alternate data models [1,2]. The primary challenges in adopting event based models relate to syntactical issues with the representation of questionnaire data. This was addressed through time-consuming clerical work, which is likely to be unsustainable over the whole of the ALSPAC data set or in other studies. In future, alternative coding schemes – such as MESH, HASSET or the Data Documentation Initiative (DDI) formatted cohort data dictionary developed by the CLOSER cohort consortium (http://www.closer.ac.uk/data-resources/closer-search-platform/) may be more efficient if they are found to fully combine clinical and inter-disciplinary survey domains. Other problems relate to missing clinical measurement units, which can be resolved by changing the EPR extraction protocols. Despite some initial problems, our work to date suggests the validity of this data warehouse design. Future work will integrate a wider range of data (ALSPAC clinical assessment data, genetic data) to complete our use case data set and then use the data warehouse to conduct exemplar epidemiological investigations using our asthma use case.

Acknowledgements

We are extremely grateful to the ALSPAC families, recruiting midwives and the ALSPAC team. The UK Medical Research Council, Wellcome Trust (Grant ref: 102215/2/13/2) and the University of Bristol provide core support for ALSPAC. The authors will serve as guarantors for the contents of this paper. This research was specifically funded by The Wellcome Trust (Grant ref: WT086118/Z/08/Z).

References

[1] Denazas D, George J, Herrett, et al. Data Resource Profile: Cardiovascular disease research using linked bespoke studies and electronic health records (CALIBER). *Int J Epidemiol* (2012) 41 (6): 1625-1638.
[2] Ford DV, Jones KH, Verplancke J-P, et al. The SAIL Databank: building a national architecture for e-health research and evaluation. *BMC Health Serv Res* (2009) 9:157.
[3] Boyd A, Golding J, Macleod J, et al. Cohort profile: the 'children of the 90s'—the index offspring of the Avon Longitudinal Study of Parents and Children. *Int J Epidemiol*. 2012 Apr 16:dys064.
[4] Burton PR, Murtagh MJ, Boyd A, et al. Data Safe Havens in health research and healthcare. *Bioinformatics*. 2015;31(20):3241–8. doi:10.1093/bioinformatics/btv279.
[5] Bache R, Daniel C, James J, et al. An Approach for Utilizing Clinical Statements in HL7 RIM to Evaluate Eligibility Criteria, accepted for 25th European Medical Informatics Conference - MIE2014.
[6] ISO 21090:2011 Health informatics -- Harmonized data types for information interchange, International Standards Organisation (ISO), 2011.
[7] Karmen C, Ganzinger M, Kohl CD, et al. A framework for integrating heterogeneous clinical data for a disease area into a central data warehouse. *Stud Health Technol Inform*. 2014;205 1060-1064.

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-161

Combining Different Privacy-Preserving Record Linkage Methods for Hospital Admission Data

Jürgen STAUSBERG[a,1], Andreas WALDENBURGER[a], Christian BORGS[b], Rainer SCHNELL[b]
[a] *Institute for Medical Informatics, Biometry and Epidemiology, Faculty of Medicine, University Duisburg-Essen, Essen, Germany*
[b] *German Record Linkage Center, University Duisburg-Essen, Duisburg, Germany*

Abstract. Record linkage (RL) is the process of identifying pairs of records that correspond to the same entity, for example the same patient. The basic approach assigns to each pair of records a similarity weight, and then determines a certain threshold, above which the two records are considered to be a match. Three different RL methods were applied under privacy-preserving conditions on hospital admission data: deterministic RL (DRL), probabilistic RL (PRL), and Bloom filters. The patient characteristics like names were one-way encrypted (DRL, PRL) or transformed to a cryptographic longterm key (Bloom filters). Based on one year of hospital admissions, the data set was split randomly in 30 thousand new and 1,5 million known patients. With the combination of the three RL-methods, a positive predictive value of 83 % (95 %-confidence interval 65 %-94 %) was attained. Thus, the application of the presented combination of RL-methods seem to be suited for other applications of population-based research.

Keywords. Duplicates, health services research, hospital, record linkage

1. Introduction

Record linkage (RL) is a common task in health care. Usually, new information is linked to an already existing record by clerical edit or by file matching using a unique identifier such as a social security number. However, clerical edit is inefficient for large data sets; one-to-one identifiers may be missing or may contain errors. Furthermore, due to data privacy, the use of unencrypted person characteristics or identifiers could be restricted. Under such constraints, RL could be performed "privacy-preserving" [1]. Typically, privacy-preserving RL is based on one-way encrypted person characteristics, either as a deterministic (DRL) or probabilistic approach (PRL). Some drawbacks of DRL and PRL could be avoided by a recently proposed RL-method using Bloom filters [2]. Bloom filters allow the computation of identifier similarities while preserving the privacy of the encrypted data. Currently, it is not clear, which RL-methods should be used for what kind of specific purpose. Therefore, this project aimed at a combination

[1] Corresponding author, Institute for Medical Informatics, Biometry and Epidemiology, Faculty of Medicine, University Duisburg-Essen, Hufelandstrasse 55, 45122 Essen, Germany; E-mail: stausberg@ekmed.de.

of DRL, PRL and Bloom filters for hospital admission due to the high importance of avoiding duplicates in hospital information systems. Although this application does not require a privacy-preserving approach, a preselection of possible duplicates would facilitate manual corrections beyond the admission unit.

The project was set up in two stages. The three RL-methods were adapted to the use case hospital admission in a development stage [3]. In this paper, the results of the validation stage is reported. On one hand, the aim was to give recommendations for the application of the different RL-methods in the use case of hospital admission. On the other hand, experiences should be gained with a large real world data set in health care that could be transferred to other scenarios as well.

2. Methods

2.1. Preprocessing

The data set used for this work included patients from the University Hospital of Munich from 1995-2014. Each patient's record contained 12 variables such as family name and date of birth. All variables had missing values. In total, 3,174,240 records were available in the data set. The data set included records that do not correspond to an identifiable natural person, for example, the record has been entered for experimental studies ("dummy patients") or are unidentified emergency patients.

For DRL and PRL, the records were standardized in order to catch common errors or ambiguities. This process was based on the UNICON-approach (cf. [4]). Common standardization steps were to replace all non-German accented characters by their unaccented replacements, removal of special symbols, converting to uppercase, and normalizing German characters. Names (family, given, and birth names) were split into 3 components, where common affixes were collected into the third field. For each name, the corresponding phonetic encoding was created [5]. In total, the 12 variables were mapped into 25 fields. Finally, the records were anonymized by using the hash values generated with the SHA-256 algorithm (along with a random salt string). Bloom filters were independently computed for all records of the data set. A newly created record identifier (ID) guaranteed the comparison between DRL, PRL and Bloom filters. For validation purposes, the data owner preserved a mapping table from the ID to the originally patient number. The 3,174,240 records were randomly split into two sets of 1,587,120 records, one for the development stage, and one for the validation stage. During development, parameters for all used methods were estimated. Based on these estimates, the procedures were applied to the validation sample. To evaluate discrepancies between the validation sample and the results of the proposed procedures, a small sample of records was checked manually.

2.2. Deterministic and probabilistic record linkage

The most straight forward way to compare records is to simply take a pair of records and compare it field by field. Since we had to work with hash values, similarities between field values are meaningless. Based on the development stage, records were identified as match if A) they shared 23 field values or B) the values agreed in the fields phonetic first name, phonetic last name, first component of birth name, date of birth and post code.

A version of the algorithm developed by Fellegi and Sunter [6] was used for PRL. The idea is to determine the partial weight of a field by considering both the frequency of the values in the underlying data set and the probability that a value may contain errors. The partial weights for the fields are then summed up to yield the weight for the record pair.

In our application, missing values were treated specially. If at least one field was empty, the partial weight was zero. An extension to this algorithm was also used here ("array matching"). Groups of fields, as the three components derived from the family name, were not only compared against their immediate counterparts but also against other attributes of that group. This was done in order to catch instances where names had been swapped, such as the name "Theresa-Marie" having been recorded once as "Theresa-Marie" and another time as "Marie". The threshold between matches and non-matches was determined by a semi-supervised algorithm called Classification for Record Linkage with Artificial Training Sets (CLARA) [7]. In the development stage, CLARA calculated a weight of 34,897 ± 0,700 as threshold.

2.3. Bloom Filters

Two different sets of variables were used. Set A includes first name, last name, birth name, sex, date of birth, place of birth, and nationality, set B additionally post code, place of residence, title, and name supplements. For both sets, the cryptographic longterm key was set up with a length of 1,000 bits and 10 hash functions. Matches were defined as pairs with a Tanimoto similarity of 1 for set A and 0.95 and higher for set B, as determined by the Multibit tree algorithm [8].

2.4. Adaptation to the Use Case Hospital Admission

The validation should mirror one year of inpatient hospital admissions. Assuming 100 daily admissions with 80 % initially as new identified patients, a 2 % sample of records was randomly extracted from the validation data set. Those records represented new patients. The remaining records represented known patients. After the exclusion of a few suspecting records, the sample of new patients comprised 31,742 records, the sample of known patients included 1,554,732 records. Blocking was applied based on the fields phonetic first name, phonetic last name and year of birth.

2.5. Validation

A pair of two records, one from the data set of new patients and one from the data set of known patients, was considered a match if it exceeds the threshold of the specific RL-method. The other pairs were considered as non-matches. For the validation of the results, a sample size calculation gave a minimum of 30 pairs necessary to distinguish between positive predictive values (PPVs) of 50 % and 25 % with a power of 80 % and an alpha of 0.05. Staff from the University Hospital of Munich consulted the hospital information system and rated a match as "true", "twin", "false", "unclear" and "dummy". From these results, the PPV was calculated based on the optimal combination of the different approaches.

3. Results

After blocking, 6,570 potential pairs remained (cf. figure 1). The application of CLARA in the development data set resulted in a slightly lower weight of 34,642 ± 0,730 leading to a threshold of 33,912. DRL found 69 matches, one of them solely identified using rule B (see above, PRL-weight was 60,207). Nine out of 69 matches (13 %) had a weight below the CLARA-threshold for the PRL (minimum 27,768). PRL gave 156 matches (weight maximum 95,635). With Bloom filters, 115 matches were found (PRL weight between -5.988 and 83,795), 43 already identified by PRL and 39 already identified with DRL. Thirty-three matches were identified by all three RL-methods.

For evaluation purposes we selected a non-random sample of 15 PRL matches closest to the CLARA-threshold, the DRL match identified by rule B, and 14 PRL non-matches closest to the CLARA-threshold either identified as match using DRL or identified as match using Bloom filters. From the 30 Matches, nineteen were rated as true duplicates, six as dummy patients, and five as unclear. Neither twins nor false positives were observed. For the PPV-calculation we rated dummy patients as true matches. Then, the PPV was 83 % (25 true matches out of 30 matches, 95 %-confidence interval 65 %-94 %).

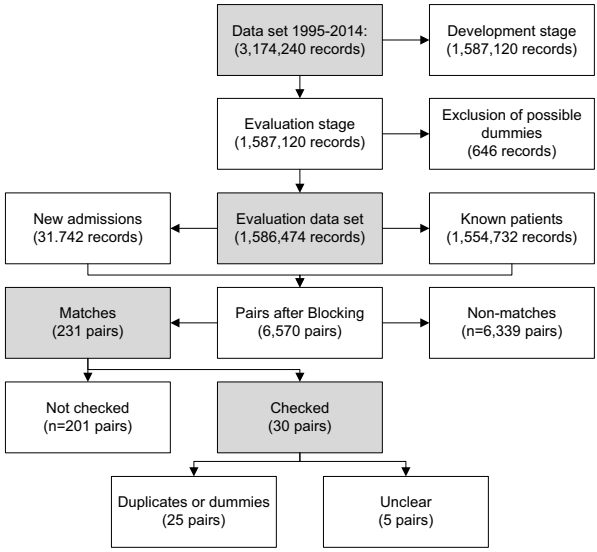


Figure 1. Flow chart of the data selection and RL steps.

4. Discussion

With the combination of the three RL-methods, we achieved a PPV of 83 % confirming the result from the development stage for DRL. This PPV is quite satisfactory for the use case at hand, the detection of duplicates at hospital admission. For that use case, the PRL-threshold could be further lowered to minimize the number of false negatives. The combination of the three RL-methods could be applied each night for patients admitted as new ones. Then, the matches would be manually inspected the next day. Based on

this study, we now estimate an amount of 5 per thousand duplicates in the data base. Therefore, even after carefully checking all admissions with the support of a hospital information systems, around 400 duplicates out of 80.000 inpatient admissions will occur annually in our study hospital.

Moreover, even with clear text and using all available information, it was not possible to distinguish between true and false positives in 17 % of the matches (5 out of 30 matches). During hospital admission, such cases can be avoided by using context information or by collecting more information from the patients. However, without those options, the risk of linkage errors by matching unrelated records is relevant in clinical practice. Since the clinical consequences of such linkage errors may be unbearable, the application of RL-methods for clinical work cannot be recommended yet. However, the application of the presented combination of the RL-methods seems to be suited in population-based research (cf. [9, 10] as examples).

Some research questions remain. Detecting the threshold in DRL and PRL still remains a major issue. Unfortunately, it was not possible to increase the validation sample due to lack in resources. The discrepancies between DRL and PRL on one hand and Bloom filters on the other hand might be due to different loss functions used for the determination of matches: Avoiding false positives (DRL, PRL) or avoiding false negatives (Bloom filter).

Acknowledgements

The project is funded under reference number STA 454/16-1 from the Deutsche Forschungsgemeinschaft (DFG). We would like to thank the University Hospital of Munich for the provision of the anonymized data set.

References

[1] P. Christen, Data Matching. *Concepts and techniques for record linkage, entity resolution and duplicate detection*, Springer, Heidelberg, 2012.
[2] R. Schnell, T. Bachteler, J. Reiher, Privacy-preserving record linkage using Bloom filters, *BMC Medical Informatics and Decision Making* **9** (2009), 41.
[3] A. Waldenburger, D. Nasseh, J. Stausberg, Detecting duplicates at hospital admission: comparison of deterministic and probabilistic record linkage, *Stud Health Technol Inform* **216** (2016), 135-138.
[4] H. Hinrichs, *Bundesweite Einführung eines einheitlichen Record Linkage Verfahrens in den Krebsregistern der Bundesländer nach dem KRG, Abschlussbericht, Projekt Deutsche Krebshilfe. Antragsnummer 70-2043-Ap I*, OFFIS, Oldenburg, 1999.
[5] H.-J Postel, Die Kölner Phonetik – Ein Verfahren zur Identifizierung von Personennamen auf der Grundlage der Gestaltanalyse, *IBM-Nachrichten* **19** (1969), 925-931.
[6] I. Fellegi, A. Sunter, A theory for record linkage, *Journal of the American Statistical Association* **64** (1969), 1183-1210.
[7] D. Nasseh, J. Stausberg, Evaluation of a binary semi-supervised classification technique for probabilistic record linkage applications, *Methods of Information in Medicine* **55** (2016), 136-143.
[8] R. Schnell, Privacy-preserving record linkage and privacy-preserving blocking for large files with cryptographic keys using multibit trees, *Proceedings of the Joint Statistical Meetings*, American Statistical Association, 187-194.
[9] F. Borst, F.A. Allaert, C. Quantin, The Swiss solution for anonymously chaining patient files, *Stud Health Technol Inform* **84** (2001), 1239-1241.
[10] B. Kijsanayotin, S.M. Speedie, D.P. Connelly, Linking patients' records across organizations while maintaining anonymity, *AMIA Annu Symp Proc* (2007), 1008.

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-166

Application of Correspondence Analysis to Graphically Investigate Associations Between Foods and Eating Locations

Andrew N. CHAPMAN[a,] , Eric J. BEH[c], Luigi PALLA[a,b1]
[a] *Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine*
[b]*Farr Institute of Health Informatics,University College London*
[c]*School of Mathematical & Physical Sciences, University of Newcastle, Australia*

Abstract. This paper presents the application of correspondence analysis (CA) for investigating associations using confidence regions (CRs) with a focus on facilitating mining the data and hypothesis generation. We study the relationship between locations and "less-healthy" food consumption by UK teenagers. CA allows for a quick visual inspection of the various association structures that exist between the categories of cross-classified variables in large datasets derived with varying study designs. The hypotheses generated by the visual display can then be independently tested using suitable regression models. CA makes use of readily available software tools and of robust statistical tests amenable to interpretation.

Keywords. Correspondence Analysis, Confidence Regions, Hypothesis generation, Hypothesis testing, clustered data, hierarchic data, location, food-group, healthy.

1. Introduction

Effective public health policies are needed to discourage less-healthy eating habits. Identifying where such habits develop may facilitate any intervention to modify them. Recently, the location of eating has been related to change in social context [1] and also linked to diet quality in adolescents [2]. In order to investigate the latter issue further, we used the National Diet and Nutrition Survey (NDNS) database for the years 2008-2011 to analyse the association between eating locations and "less-healthy" food-groups consumed by UK teenagers aged 11-18 years. The published description of the NDNS programme [3] is used by the federal Government to monitor progress on diet and nutrition objectives. We focused on teenagers as habits taken up at that stage may have an effect throughout life and thus are potential intervention targets.

The response rate for the completion of a four-day food diary and a lifestyle interview was 56%. A total of 884 teenagers aged 11 to 18 years responded, providing a total of 62,523 diary records. The mean (standard deviation) of number of diary entries over four days for an individual teenager is 71 (22). Such entries are not independent of each other due to a two-level hierarchy of correlation in the diary records: individual and meal-time. Our analysis was performed purely on instances

[1] Corresponding author, luigi.palla@lshtm.ac.uk .

rather than quantities and addressed the open question whether there is a relationship between less-healthy eating and the location where the food is consumed.

2. Methods

We randomly selected half the NDNS data diary entries for hypothesis generation by cross-tabulating foods and eating locations, and the remaining half of the diary entries were used for performing the hypothesis tests that we generated.

2.1. Correspondence Analysis CA and Confidence Regions (CRs)

The contingency table (frequency matrix) was analysed by Correspondence Analysis (CA) (cf [4] for practical guide and [5] for a historical and up-to-date theoretical treatment of CA). CA allows us to investigate the relationship between the categories of two variables, by projecting them on two dimensions (axes) that jointly represent a large percentage (displayed on the axes) of the χ^2 deviation (called inertia) of the table from the condition of independence between the two variables (foods and locations).

There are two ways to plot CA results in the two most informative dimensions: symmetric plots and biplots. In symmetric CA plots, both row and column profile-points are normalised. This is useful for row-row and column-column associations but distorts [4] associations between rows and columns. Hence we preferred the biplot where row profiles are normalised (rescaled) but column profiles are not (or vice versa) and the direction of the column profiles are shown as arrows from the origin to facilitate the interpretation of row-column associations via the size of the angle between the categories. We can also sometimes find meaning for an axis by considering the opposition and placement of points along that axis. Importantly, this descriptive method does not require any assumptions about the data. Therefore, it can be used to explore correlated data. We also verified that the complex survey design should not affect the CA [7] and the application of survey weights in the analysis yielded negligible differences in the results/plots.

Two algorithms using the freely available "R" software routines are available to enhance CA plots with elliptical confidence regions (CRs) for each profile-point: CAvariants uses algebra and parametric-assumptions [8] and CABOOTCRS uses a bootstrap method [6]. We used CRs from CABOOTCRS with a 95% confidence level to eliminate those locations and food-groups with larger CRs that overlap the origin, since their deviation from the average profile could be explained by sampling variation. Associations amongst the remaining food-groups and locations suggested hypotheses which were tested using logistic regression/Generalized Estimating Equations (GEE).

Our CA plots for all food-groups were cluttered with too many overlapping profile points and CR ellipses. So, when constructing the CRs, we sub-divided the food-groups into categories: "healthy", "neutral" and "less-healthy" as defined in [9] using the UK Food Standards Agency (FSA) nutrient profiling system detailed in [10].

The CA method allows profiles to be sub-divided without introducing inconsistencies: this is one reason for using "inertia" as the metric in CA analysis [4].

2.2. GEE and Odds Ratios ORs

For hypothesis testing, we accounted for correlation by performing logistic regression by Generalised Estimating Equations (GEEs) yielding population mean ORs. GEEs provides unbiased estimates [11] of ORs despite our ignorance of the true correlation of diary entries within mealtimes, and within/between individuals. Since the correlation structure in the model is unreliable, the GEE model (via the SAS procedure GENMOD) provides empirical estimates of standard errors.

3. Results

Using a random process to split the diary records for the top 25 food-groups resulted in a hypothesis generating dataset of 20,567 diary records and a testing dataset of 20,455.

We plotted twelve of the top 25 food-groups which were classified as less-healthy: Biscuits, Crisps, Chocolate and Sweets, Buns & Cakes, Miscellaneous Foods, Cheese, Low-fibre Cereals, Sausages, Fried Chicken, Lower-fat spreads, Meat-pies & Pastries, Jams. Figure 1 and figure 2 are CA plots that summarise various associations. For both figures, the first axes reflects the greatest correlation (66.82%) between foods and location while the second axis accounts for 21.4% of this association. Therefore figures 1 and 2 reflect 21.4%+66.82% = 88.22% of the association between the two variables and so provides an excellent visual summary of how the two variables are related.

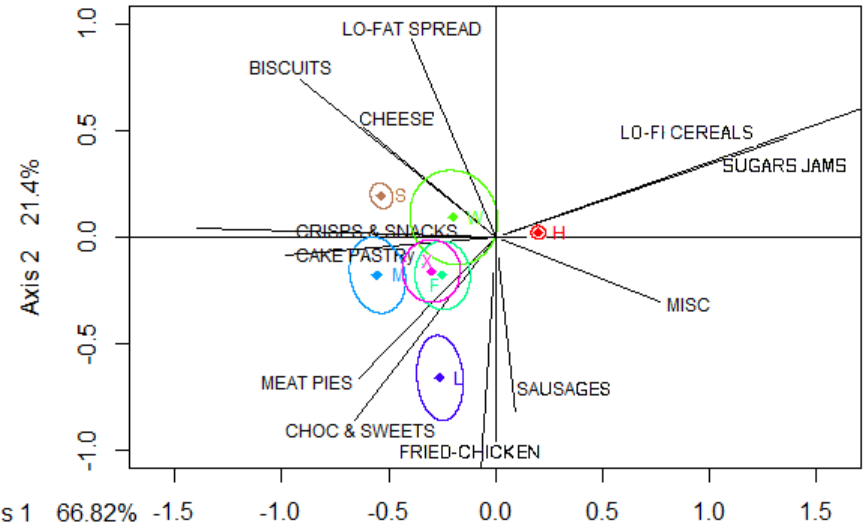


Figure 1. Biplots of Less-Healthy Foods with 95% bootstrap CRs for **Locations** H-Home S-School W-Work F-Friends/Carers L-Leisure M-Mobile X-Other

In the top-left quadrant of figure 1, Cheese, Less-fat spreads and Biscuits appear associated with School and Work. Moving anti-clockwise, we find Crisps, Cakes and Pastries are associated with non-Home locations. At the bottom of figure 2, we find Fried Chicken is associated with Leisure locations.

In both figures, Meat-Pies and Chocolate appear in the bottom-left quadrant associated with Leisure, Friend's & Carer's homes, Other and Mobile locations. In

figure 2, the "Chocolate & Sweets" CR appears entirely inside the CR for Meat-Pies; this invited further investigation.

In figure 2, the locations have aligned themselves in three main directions which we have used to simplify our hypotheses: Home, then School and Work together, and then all Other locations (Leisure, Mobile, Other, Friend's & Carer's homes). As it is usually harder to collect information on eating behaviour at "Other" locations, we focus on those food-groups associated with eating locations away from Home and away from School among which Meat-Pies and "Choc and Sweets" are the main candidates, based on the biplots.

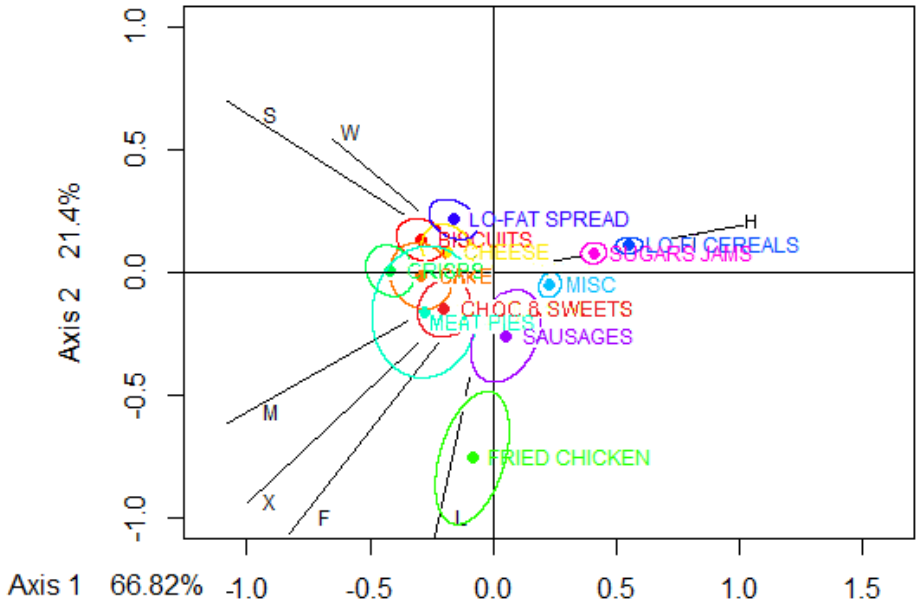


Figure 2. Biplots of Less-Healthy Foods with 95% bootstrap CRs for **Food_Groups** H-Home S-School W-Work F-Friends/Carers L-Leisure M-Mobile X-Other

Based on the findings from the two CA plots, we tested two hypotheses using logistic regression (adjusted by sex, age, weekend and socio-economic status): that "Choc and Sweets" and Meat-Pies were more likely to be consumed when a teenager finds himself/herself at "Other" locations away from School/Work and away from Home. Our results are summarised in Table 1:

Table 1. Adjusted Odds Ratio estimates for food-groups as outcomes and location-types as exposures.

Food-Group	OR vs Home	99% CI p-value	OR vs School-Work	99% CI p-value
Choc and Sweets	**2.5**	(1.8, 3.4) p<0.0001	**1.8**	(1.2, 2.8) p=0.0002
Meat Pies	**2.8**	(1.5, 5.0) p<0.0001	1.3	(0.6, 3.0) p=0.44

We also detected strong evidence (p=0.002) of a linear trend for age so that each additional year of a teenager's age increases the odds of his/her eating a Meat-Pie by a factor of 1.15 with 99% CI (1.02, 1.30).

4. Discussion and Conclusion

In this paper, we showed how correspondence analysis with confidence regions facilitated the systematic investigation of the relationship between categorical variables from a large dataset. This analysis generated plausible hypotheses concerning associations which were then tested using regression methods that take into account the correlation between observations.

Our specific aim was to explore and test for associations between foods consumed and eating locations, recorded in NDNS diet diaries. We focussed on less-healthy foods and found evidence of higher odds of consumption of those at locations away from Home and School. Therefore, public health interventions are warranted to reduce the consumption of foods like meat-pies and chocolate and sweet snacks in other locations.

Some methodological challenges still arise from our work. For example, the CRs in CA pertain to statistical significance of each of the categories of the variables and not to any measure of the association between the categories. Moreover, while the CRs were the key data-mining tool, it is plausible that the correlation in the data warrants further investigation to refine CA and the CRs. Another fundamental issue that may affect the results here is the possible bias due to under-reporting should the probability to report eating a food vary according to eating location. However, evidence of under-reporting for diet diaries is likely related to amount rather than to reporting vs non reporting, which we have overcome by counting food instances (0/1) rather than summing food quantity (grams) per location.

Despite these challenges, we stress the usefulness of CA and its CRs as a data mining tool for hypothesis generation for the analysis of cross-classified categorical variables.

References

[1] Holm L., et al, 2016, Changes in the social context and conduct of eating in four Nordic countries between 1997 and 2012: *Appetite* 103, p. 358-368.
[2] Williams J.L., 2016, Spaces between home and school: The effect of eating location on adolescent nutrition: *Ecology of Food and Nutrition*, v.55-1, p.65-86.
[3] www.gov.uk/government/uploads/system/uploads/attachment_data/file/310995/NDNS_Y1_to_4_UK_report.pdf , 2013, NDNS Report Y1 to Y4.
[4] Greenacre M. J., 1993, *Correspondence Analysis in Practice*: London, Academic Press, 195 p.
[5] Beh E. J., Lombardo R., 2014, *Correspondence Analysis: Theory, Practice and New Strategies*, Wiley.
[6] Ringrose T. J., 2012, Bootstrap confidence regions for correspondence analysis: *Journal of Statistical Computation and Simulation*, v. 82, p. 1397-1413.
[7] Nyfjall M., 2002, Aspects on Correspondence Analysis Plots under Complex Survey Sampling Designs, Research Report, Department of Information Science, Division of Statistics, Uppsala University.
[8] Beh E. J., Lombardo R., 2015, Confidence Regions and Approximate p-values for Classical and Non Symmetric Correspondence Analysis: *Commun.. in Statistics-Theory and Methods*, v. 44, p. 95-114.
[9] Pechey R., et al. 2013, Socioeconomic differences in purchases of more vs. less healthy foods and beverages: Analysis of over 25,000 British households in 2010: *Social Science & Medicine*, v. 92, p. 22-26.
[10] www.food.gov.uk, 2011, Nutrient Profiling Technical Guidance, in F. F. S. Agency
[11] Liang K. Y., Zeger S.L., 1986, Longitudinal data-analysis using generalized linear-models: *Biometrika*, v. 73, p. 13-22.

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-171

Data Driven Quality Improvement of Health Professions Education: Design and Development of CLUE – An Interactive Curriculum Data Visualization Tool

Claire Ann CANNING[a,1], Alan LOE[a], Kathryn Jane Cockett[a], Paul GAGNON[a], Nabil ZARY[b]
[a]*Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore*
[b]*Department of LIME, Karolinska Institutet, Stockholm, Sweden*

Abstract. Curriculum Mapping and dynamic visualization is quickly becoming an integral aspect of quality improvement in support of innovations which drive curriculum quality assurance processes in medical education. CLUE (**C**urricu**lu**m **E**xplorer) a highly interactive, engaging and independent platform was developed to support curriculum transparency, enhance student engagement, and enable granular search and display. Reflecting a design based approach to meet the needs of the school's varied stakeholders, CLUE employs an iterative and reflective approach to drive the evolution of its platform, as it seeks to accommodate the ever-changing needs of our stakeholders in the fast pace world of medicine and medical education today. CLUE exists independent of institutional systems and in this way, is uniquely positioned to deliver a data driven quality improvement resource, easily adaptable for use by any member of our health care professions.

Keywords. Curriculum mapping, visualization, design based approach, online interactive map, searchability functions

1. Introduction

In Medical Education, and increasingly in higher Education in general, curriculum mapping has become an essential component of curriculum integrity, serving many purposes for multiple stakeholders. For example, it enables (i) learners to dynamically track their progress and thus promote transparency of the taught curriculum, (ii) educators to maintain perspective with respect to the alignment of the curriculum with intended outcomes, and (iii) administrators to more effectively manage the quality assurance mandate [1]. There are few examples in the literature of detailed curriculum mapping tools, despite their prevalence in use in Medical Education. One example is LOOOP, a curriculum mapping tool that was developed to map, plan and accredit 22 competency based courses across Europe, Africa and Asia [2]. This mapping process was developed using 4 out of the 10 windows described by Harden [1], namely outcomes, content, learning opportunities and assessment. Although some curriculum mapping

[1] Corresponding Author, Claire Ann Canning, Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore 308232; Email: ClaireAnnCanning@ntu.edu.sg

tools are in use in many medical schools, they don't exist independent from underlying institutional technological environments which limits their use in other contexts. Furthermore, the data is not structured in a manner that complies with existing technical standards which therefore doesn't allow the sharing and benchmarking of curricula across different curriculum mapping systems. The aim of this project is thereby to design and develop a new bespoke curriculum mapping system that addresses both stated shortcomings. The targeted system is named CLUE.

2. Methods

2.1 Mapping the Curriculum

A bespoke outcomes based curriculum was developed for a systems based undergraduate MBBS programme. Learning outcomes were developed (along with the relevant learning material) which were aligned to the UK's General Medical Council (GMC), the Singapore Medical Council (SMC) and the Accreditation Council for Graduate Medical Education (ACGME). The day to day learning opportunities along with specific learning outcomes and learning resources over two years were developed *de novo* and captured in the iLKC (Compass) suite, allowing students and faculty alike to navigate a digital learning environment throughout the 5 years of their undergraduate curriculum [3]. For the purpose of this report we are focusing on the first two years of the undergraduate MBBS curriculum. No previous curriculum mapping system was in place, and the intended use is for research, curriculum development and quality improvement purposes. The stakeholders currently utilizing this system are medical education researchers, faculty, students and senior management. The curriculum is mapped to three overarching themes which span across 5 years (see Figure 2), with greater emphasis on each theme as the student's progress through the appropriate stage of learning and training, Theme 1: Scientific Basis of Medicine, Theme 2: Clinical Management and Patient Centred Care, Theme 3: Healthcare Delivery and Professional Standards. Underpinning these themes, especially in the early years, are horizontal and vertical courses focusing primarily on Integrated Science in a Medical Context. A course such as Human Structure and Function spans the majority of year 1 and year 2 and is comprised of blocks of time (teaching blocks, a total of 8 across the first two years) devoted to human body systems, e.g. the Cardiorespiratory system. The Cardiorespiratory block is subsequently divided into various topics and so on. Each individual topic e.g. Haemorrhage, has multiple learning outcomes related to the curriculum. Each learning event is mapped to the fundamental units of our curriculum, e.g. anatomy, communication, disease investigation. In addition, each learning event is mapped to an appropriate mode of assessment. The final layer of mapping is the alignment of learning opportunities to governing frameworks, e.g. GMC, SMC and ACGME, to ensure our taught curriculum maps to desired graduate outcomes and competencies.

2.2 Approach to Design and Develop the Curriculum Mapping System

To ensure the most effective design and development process a design based approach was adopted [4]. Stakeholders were encouraged to provide regular input on how best to (i) visualise the data and (ii) deliver the required functionality. Using this interactive and

iterative process, our current curriculum mapping tool CLUE, or Curriculum Explorer, emerged. CLUE draws upon the curriculum materials already defined by our school**,** existing as a mapped dataset within iLKC. The suite supports the curriculum management of learning outcomes, targeted learning resources and timetabling of associated learning activities within each teaching block across the years. Each of these components has its own designated database (See Figure 1). Data from each of these databases is transferred and stored in a data warehouse. The curriculum mapping tool (CLUE) then draws the transformed data from the data warehouse, and enables subsequent tracking of changes to outcomes, resources etc.

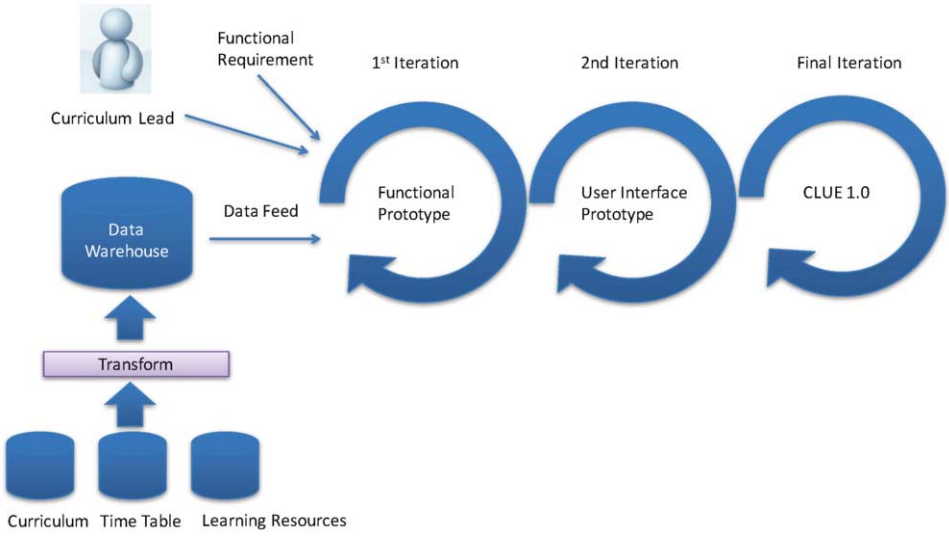


Figure 1. Iterative systematic design of CLUE.

Figure 1. Curriculum, Timetable and Learning Resources data is transformed and migrated to the data warehouse. Curriculum and Technical Leads devise the functional requirements. The data are fed into the first iteration, a function prototype, which is modified to become the user interface prototype (2nd iteration) which ultimately culminated in the final iteration, CLUE.

3. Results

3.1 CLUE Visual Explorer

Here we describe the generation of CLUE, focusing on the first two years of an undergraduate MBBS programme, and we describe the searchability and mapping functions of this dynamic and interactive resource. CLUE can be visualized via multiple paths depending on the level of the curriculum being interrogated. One main path is via an inherited mode, where one navigates through the curriculum starting at the top of the

hierarchy, i.e. Theme level, and proceed through the underlying Fundamentals and relevant Domains, as demonstrated in Figure 2 A. The same type of search can be completed in a non-inherited mode simply by clicking on the desired domain, e.g., Anatomy. Either approach will highlight all areas in the curriculum, and all learning events where Anatomy is taught across the years. The user can then review the learning outcomes and learning resources specific to the Anatomy topics highlighted, as well as the mode of assessment relevant to that learning activity. A complementary approach to visualising the curriculum is to search by year, by course, or teaching block etc. and to manually trace through the curriculum in a hierarchical manner (Figure 2 B). Alternatively, a *free text search* function allows the user to narrow in on any topic of interest, e.g. Blood Gases (Figure 2 C).

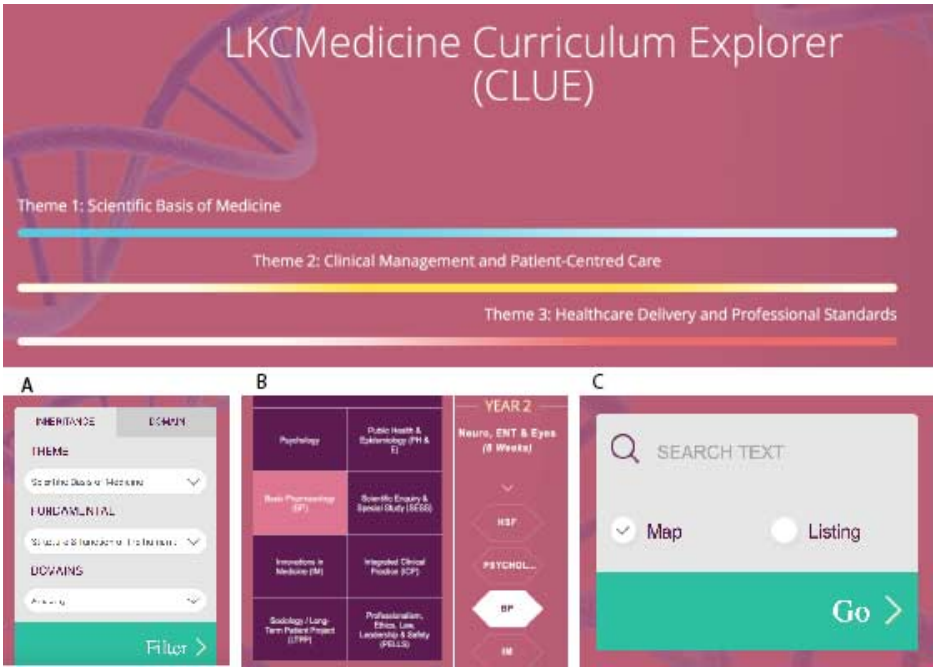


Figure 2. The various exploration paths of CLUE

Figure 2. The curriculum Themes, Fundamentals and Domains can be searched in a hierarchical manner (Figure 2A). Alternatively, by searching by year (e.g. year 2) one can search the courses taught in year 2, e.g. Basic Pharmacology, and all the areas where Basic Pharmacology is taught will appear as highlighted icons (Figure 2B). Lastly, an open text search allows you to search the curriculum by free text (Figure 2C).

4. Discussion

We have developed a curriculum visualization system (CLUE) that supports future iterations tailored to our specific shareholder's needs. CLUE enables the user to navigate through each of the 5 years of our curriculum via multiple paths. For example, by

choosing a particular year, the courses taught within that year are immediately visible. In the same manner, it is also possible to view where integrated courses are taught throughout and across the years. Our curriculum is comprised of three Themes, and the curriculum mapped can then be viewed at the level of each these Themes, the underlying curriculum Fundamentals and learning Domains, in a hierarchical manner (e.g., Theme 1>Scientific Basis of Medicine > Fundamental > Structure and Function of the Human Body in Health > Domain > Anatomy). In this manner, the user steps down through the curriculum hierarchy and arrives at a granular display of all the learning outcomes that are mapped to Anatomy. Alternatively, the user can choose a course, e.g. Basic Pharmacology (BP), and see via highlighted areas where, across the years, Basic Pharmacology is taught. In an individual teaching block, e.g., the Neuro, ENT, Eye (NEE) teaching block in year 2, the user can click the BP icon and it will show all the learning outcomes related to Basic Pharmacology and relevant to the NEE teaching block. Clicking any of the displayed learning topics will lead to the full display of the learning outcomes, the relevant assessment mode, and how this topic has been mapped to curriculum Themes, Fundamentals and Domains. Finally, this learning event is also mapped to the outcomes of multiple competency frameworks, namely the GMC, the SMC and the ACGME. An additional feature of CLUE is a search function by free text, which enables the user to search the curriculum, for any free text content, e.g., if one were to search for "Blood Gases", one would be taken to all the courses across the five years where Blood Gases are taught. CLUE builds on an iterative and dynamic process to reflect and support the fast pace of curriculum change in medicine today. As the curriculum evolves, occasioned by reviews of accreditors, and modifications to competency frameworks, so too must CLUE. CLUE must be adaptable and easily modifiable if our multiple stakeholders are to use the tool to identify changes to Outcomes, associated Learning Resources and the variety of Assessments internal to the successful training and development of our trainee doctors. In this manner, and with these attributes, CLUE forms part of an innovative and exciting research approach to curriculum design development and programme evaluation in Medical Education.

References

[1] R. Harden AMEE Guide No. 21: Curriculum mapping: a tool for transparent and authentic teaching and learning. *Medical Teacher* 2001 Mar;23(2):123-137.

[2] F. Balzer, W.E. Hautz, C. Spies, A. Bietenbeck, M. Dittmar, F. Sugiharto, L. Lehmann, D. Eisenmann, F. Bubser, M. Stieg, S. Hanfler, W. Georg, A. Tekian, O. Ahlers. Development and alignment of undergraduate medical curricula in a web-based, dynamic Learning Opportunities, Objectives and Outcome Platform (LOOOP). *Medical Teacher* 2016, 38(4):369-377

[3] P. Gagnon, R. Mendoza, J. Carlstedt-Duke. *The flipped classroom: Practice and practices in higher education* In C. Reidsema, L. Kavanagh, R. Hadgraft, & N. Smith, (Eds.). Springer Singapore 2017

[4] A. Collins. Towards a design science of education. In E. Scanlon & T. O'Shea (Eds.), *New directions in educational technology*. Springer Berlin, Germany 1992

[5] T. Reeves, S. McKenney, P. Herrington. Publishing and perishing: The critical importance of educational design research. *Australasian Journal of Educational Technology* 2011, 27(1), 55-65

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-176

Developing Healthcare Data Analytics APPs with Open Data Science Tools

Bibo HAO[a,1], Wen SUN[a], Yiqin YU[a], Guotong XIE[a]
[a] *IBM Research - China, Beijing, China*

Abstract. Recent advances in big data analytics provide more flexible, efficient, and open tools for researchers to gain insight from healthcare data. Whilst many tools require researchers to develop programs with programming languages like Python, R and so on, which is not a skill set grasped by many researchers in the healthcare data analytics area. To make data science more approachable, we explored existing tools and developed a practice that can help data scientists convert existing analytics pipelines to user-friendly analytics APPs with rich interactions and features of real-time analysis. With this practice, data scientists can develop customized analytics pipelines as APPs in Jupyter Notebook and disseminate them to other researchers easily, and researchers can benefit from the shared notebook to perform analysis tasks or reproduce research results much more easily.

Keywords. Healthcare Data Analytics, Jupyter Notebook, Analytics Application

1. Introduction

Emerging theories and tools in data science benefit researchers who work on analyzing healthcare data. Among which, Jupyter Notebook is a user-friendly and open source tool that allows data scientists to develop analysis programs and capture the analysis process and results in an agile way. There have been some reference cases of using Jupyter Notebook to perform collaborative and reproducible analyses in many scientific research areas [1], [2], [3], healthcare data analytics also included [4], [5].

However, current practice requires researchers to develop programs to analysis data with programming languages like Python, R, etc. Although there are courses teaching biomedical researchers to learn these programming languages [6], for many researchers, there is still a steep learning curve for them to grasp skills of developing or understanding program source code, especially for those without a background in computer science education. Besides, when explaining the data analysis process and results to other researchers, tables and charts are better ways to communicate than program code.

How can we make data science more approachable to all researchers? In this study, we aim at develop a practice and designing a system for users to develop user-friendly and reusable data analytics APPs (applications), so that researchers can convert existing notebooks to analytics APPs with very limited efforts. The converted analytics APPs, with rich interactions and real-time analysis capabilities, can then be used to perform specific analytics tasks on different datasets. In this way, data scientists can develop

[1] Corresponding Author. Email: haobibo@cn.ibm.com

customized, reusable applications with rich interactions to make their analytics methods and solutions more approachable to all researchers.

2. Methods

2.1. Develop Analytics Pipelines using Jupyter Notebook and Open Data Science Tools

Incubated from interactive scientific program developing practice, Jupyter Notebook is a tool now widely used in areas like scientific research and education. Researchers and students benefit from Jupyter Notebook's features of elaborating data process pipelines with runnable program source codes, comments, and explanations, as well as prompt results like tables, charts, and HTML elements. The developed pipelines can then be shared to validate or reproduce previous analytics pipeline with very little efforts of configuring environments. Among tens of programming languages supported by Jupyter Notebook, Python and R are currently most popular ones in the data science community. By using open source libraries like pandas, matplotlib, scipy, scikit-learn, etc., researchers can perform tasks like data cleansing, wrangling, and visualization, as well as statistical analysis and modeling.

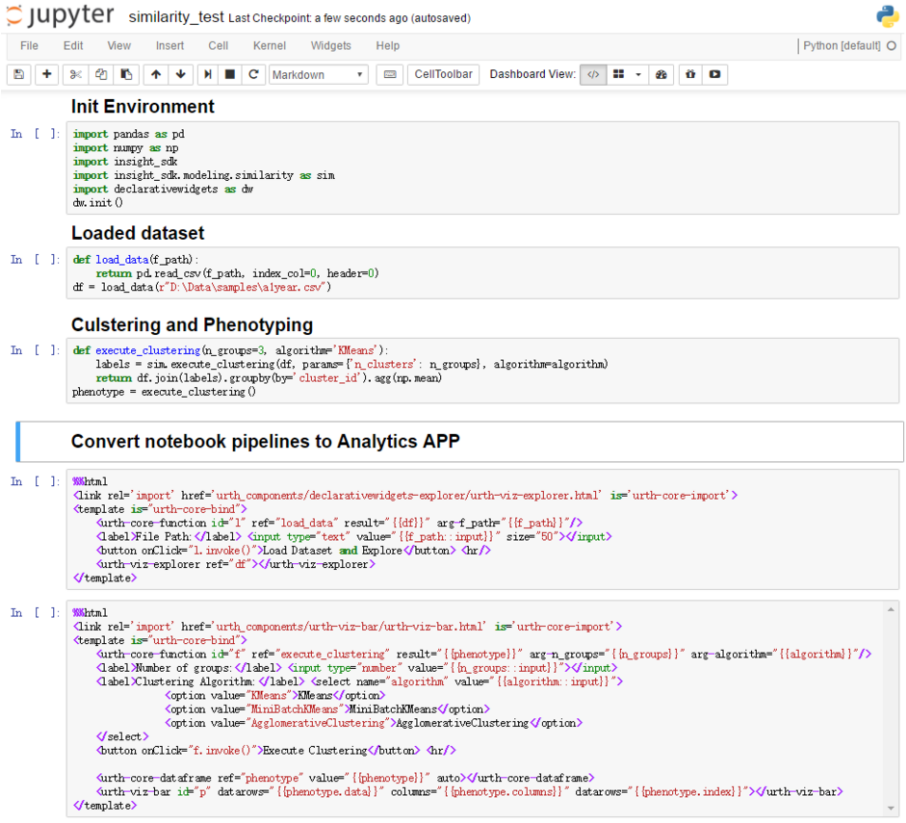


Figure 1. Screenshot of a patient similarity analysis pipeline notebook

In our practice of analyzing clinical registry data in Jupyter Notebook, we developed reusable pipelines using open source packages for data cleansing, exploration, visualization, and modeling. Users can configure some parameters to achieve specific goals, for example, try different algorithms and parameters for modeling. After setting these parameters, the pipelines can then generate analysis results presented as tables or charts for researchers. Figure 1 demonstrates a sample notebook pipeline to perform patient similarity analysis developed in Python. The process of this pipeline can be described as 1) import related packages, 2) load a dataset, 3) cluster patients into several groups and summary the patient features of each group as phenotyping (the detailed implementation of clustering algorithms is encapsulated in the "insight_sdk.modeling.similarity" Python module).

2.2. Convert Analytics Pipelines to Interactive APPs

In this part, we introduce two components, namely Jupyter Declarative Widgets [7] and Jupyter Dashboards [8], to convert notebooks into analytics APPs. The Jupyter Declarative Widgets extension provides a mechanism to bind front-end variables or HTML elements (like JSON object, input elements) to back-end data variables (like DataFrames) or functions in a live notebook, so that users can invoke code execution by clicking HTML element and get real-time results rendered into interactive HTML widgets like d3 based charts and tables. Whilst the Jupyter Dashboards provides an interface to arrange notebook cell outputs to HTML pages with self-defined layouts.

As shown in Figure 1, codes in the last two cells implement two declarative widgets via the composition of some customized HTML DOM elements provided by the Jupyter Declarative Widgets. The binding is implemented by 1) binding HTML input element attributes to function arguments, function invocation, and function returned value; 2) binding the returned value to a DataFrame variable container at front-end; 3) using a front-end template encapsulated charts or table widget to render the DataFrame variable at front-end. So once the user clicks the invocation button, the widgets will obtain parameters from front-end and send a request to back-end with parameters to execute the function. The execution result will be returned to front-end and update the HTML charts in real-time which it was bound.

Key techniques used in the analytics APP is to bind back-end variables in notebook kernel to front-end JavaScript variables. In order to implement the binding, Jupyter Declarative Widgets consist of both front-end and back-end components for two-way communication. The back-end part provides interfaces to facilitate interacting with code running on the kernel, and the front-end part sync variables with back-end through the kernel. Although we only demonstrate a Python notebook, the Jupyter Declarative Widgets also supports notebooks written in R and Scala. As a result, front-end elements can be reused as independent components even outside notebook or dashboards.

3. Results

Figure 2 shows an analytics APP converted from a notebook, where users can interact with the APP by inputting parameters and clicking buttons, and the real-time results will be displayed as charts below. Firstly, the user can input a data file path and click the button to load dataset. Then the user can preview and explore the loaded data in the below data exploration widget. After which, the user can choose a clustering algorithm and set

parameters of the number of groups in the clustering analysis. Once the clustering algorithm finished, the charts below will display summary information of patient features in each clustered group.

In this user interface, the source code is invisible, so that users can focus on interactions and the results. What's more, the visualization is bound to back-end variables, which means once the backend data is changed, the charts will update to show the latest results.

Compared to our previous practice of developing web applications for users to perform analysis tasks, this approach has significantly reduced the workload and simplifies the process of connecting the user interface to the back-end functions. When developing analytics APPs, data scientists can put analytics code written in back-end programming languages and the declarative implementation of front-end widgets in one notebook, and each notebook becomes an independent analytics APP. This method allows data scientist to leave more time on creating stronger solutions based on analytics with rich user interactions instead of using conventional web framework to connect front-end presentation and backend function executions. For scenarios like different dataset with similar schema (e.g., only one DataFrame), or streaming data, this practice helps give power to the user for rapid insights from data quickly.

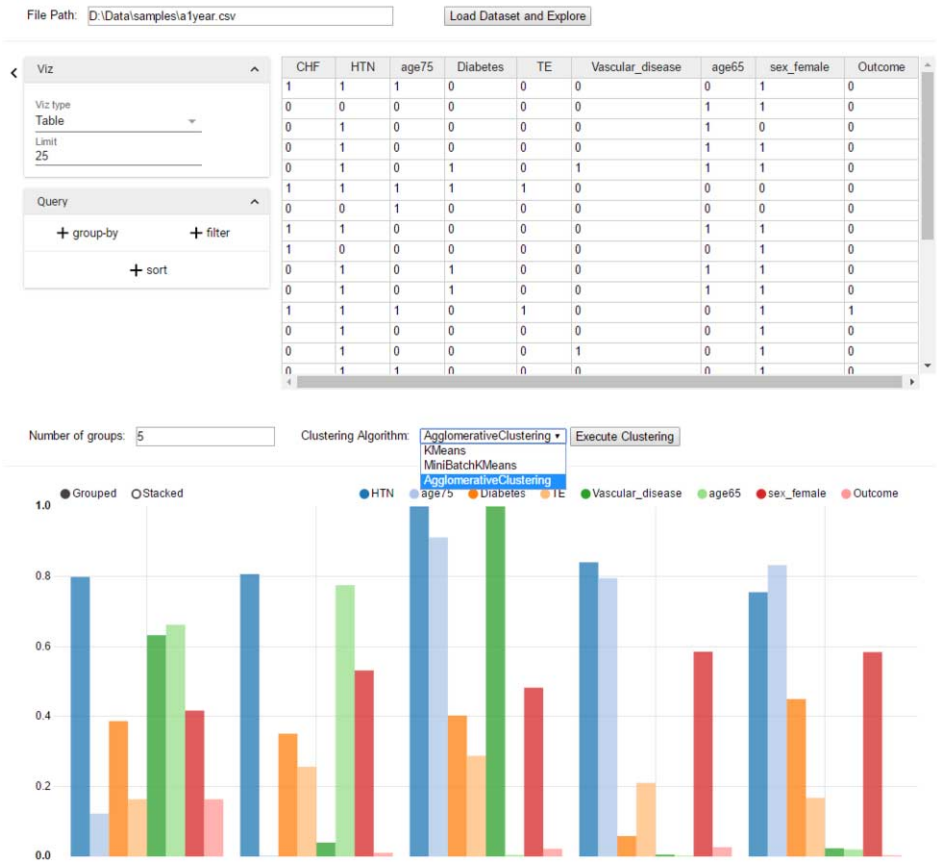


Figure 2. Screenshot of a patient similarity analysis APP converted from notebook

4. Discussion

Recent advances in big data analysis enable researchers to gain insights from healthcare data more conveniently. Among many data analytics tools, Jupyter Notebook and open source data analytics packages have proven to be a popular combination to analysis data and share insights. While developing code to conduct analysis is still not approachable for many researchers. With the help of Jupyter Widgets and Jupyter Dashboards, we have developed a practice to convert existing notebooks to user-friendly analytics APPs that allow researchers focus on interactions and analysis results. The converted analytics APP can then be used by researchers to perform customized analysis tasks with no need to read or modify the source code.

Comparing to conventional analytics tools like Weka, SPSS, SAS, etc., this practice offers both flexibility and rich interactions to users. With limited efforts, data scientists can convert their notebooks into reusable analytics APPs for researchers who cannot develop programs. This makes data science more approachable to researchers in the medical informatics areas. For example, an analytics APP to conduct statistical tests can be disseminated in the form of a Jupyter Notebook analytics APP, where end-user can use the APP by inputting several basic parameters in a web form rather than understanding and modifying the source code. In this way, researchers can focus on the analytics tasks and get rich format charts and tables.

One limitation in this practice is that for each analytics APP, the functions might be bound to some specific dataset schemas because analytics tasks are usually very ad-hoc and different datasets can vary a lot in the schemas. It is still hard to create an all-in-one APP that can handle all scenarios. Especially when handling with datasets in non-standard schemas of relational database tables, it is challenging to adapt analytics programs to different dataset schemas. One of our future research directions is trying to find a way to handle with different dataset schemas so that the analytics APPs can be adapted to different datasets more easily.

References

[1] H. Shen, Interactive notebooks: Sharing the code, Nature 515, no. 7525 (2014): 151-152.
[2] W.D. Fraser, Python-Based Data Analysis Tools for Aerospace Medical Research. Aviation, space, and environmental medicine 85, no. 5 (2014): 583-584.
[3] J.R. Stevens, M. Elver, and J.A. Bednar, An automated and reproducible workflow for running and analyzing neural simulations using Lancet and IPython Notebook, Front Neuroinform 7 (2013): 44.
[4] B. Hao, W. Sun, Y. Yu, J. Li, G. Hu, G. Xie, Accelerate Healthcare Data Analytics: An Agile Practice to Perform Collaborative and Reproducible Analyses, in MIE 2016, pp 552-556, 2016.
[5] A.K. Manrai, B.L. Wang, C.J. Patel, and I.S. Kohane, Reproducible and Shareable Quantifications of Pathogenicity, Pacific Symposium on Biocomputing Pacific Symposium on Biocomputing. 2016; 21:231-242.
[6] B.E. Chapman, J. Irwin, Python as a First Programming Language for Biomedical Scientists, in proceeding of the 14th Python in Science Conference, pp 12-17, 2015.
[7] Jupyter Declarative Widgets, [Cited 2016 Nov, 1]. Available from https://github.com/jupyter-incubator/declarativewidgets
[8] Jupyter Dashboards, [Cited 2016 Nov, 1]. Available from https://github.com/jupyter-incubator/dashboards

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-181

Fast and Efficient Feature Engineering for Multi-Cohort Analysis of EHR Data

Michal OZERY-FLATO[1], Chen YANOVER, Assaf GOTTLIEB, Omer WEISSBROD, Naama PARUSH SHEAR-YASHUV, and Yaara GOLDSCHMIDT
Healthcare Informatics Department,
IBM Research - Haifa, Israel

Abstract. We present a framework for feature engineering, tailored for longitudinal structured data, such as electronic health records (EHRs). To fast-track feature engineering and extraction, the framework combines general-use plug-in extractors, a multi-cohort management mechanism, and modular memoization. Using this framework, we rapidly extracted thousands of features from diverse and large healthcare data sources in multiple projects.

Keywords. Feature engineering, electronic health records, longitudinal data.

1. Introduction

Feature engineering is the process of deriving informative features from data for a machine learning task. It requires tailoring to the source data, and is often assisted by domain knowledge. Following the increase in size and complexity of data, feature engineering has become a much more challenging, programmatically multifaceted, and time-consuming task. We describe a framework for enhancing the process of feature engineering from EHR data. This framework is based on extensive experience in analyzing such data, gathered during more than a decade of research on multiple disease areas and data sources.

2. Feature Extraction Framework

The basic elements of the proposed framework are *feature extractors* (shortened as *extractors*). Each extractor derives a set of features from the source data, or from features computed by other extractors. The features computed by an extractor are determined by a predefined, configurable set of parameters. Commonly-used extractors and their parameters are described in Section 2.1. The dependencies between different extractors are presented by a directed acyclic graph, termed *the extractors graph*. Typically, this graph has a single root, which represents the final output features matrix. The *extraction engine* orchestrates the entire extraction process; it instantiates and manages the extractors graph, and provides additional utilities that manage: (1) access to the data source; (2) parameters; (3) cached features; and (4) train-data statistics. The

[1] Corresponding author, Healthcare Informatics Department, IBM Research - Haifa, Haifa University Campus, Haifa, 3498825, Israel; E-mail: ozery@il.ibm.com.

input to the extraction engine defines both the extractors graph and the analyzed cohort. The latter is managed in a *cohorts repository*. Figure 1 illustrates the extraction framework and its components.

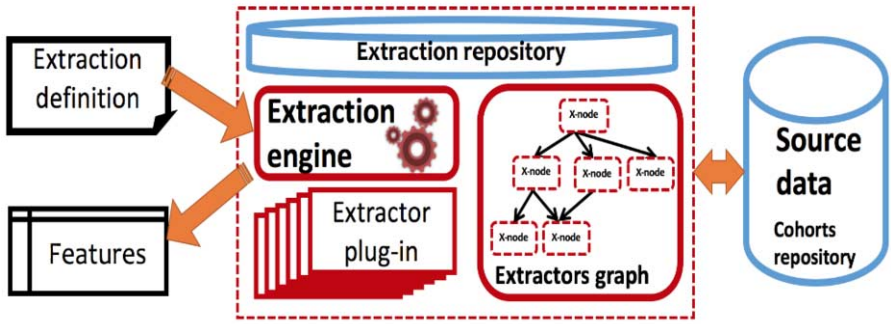


Figure 1. Overview of feature engineering framework.

2.1. Feature Extractors

Extractors are the building blocks of the feature extraction process. These plug-ins allow users to add new extractors without requiring a code change of the framework. Below we describe commonly used extractors, starting with those accessing a structured data source directly:

1. **Event extractor**: computes various aggregations on a sequence of events, where each event has a patient identifier, event name, time point/interval, and a value. A similar notion was previously described in [1].
2. **Numerical and Categorical Table extractor**: computes various aggregations on numerical and categorical data columns in a table, respectively. The input table has the following columns: patient identifier, time [optional], and one or more data columns.
3. **Time-to-Index-Date extractor**: computes various aggregations on the difference between date-columns in a table and the patient's index-date (index-dates will be discussed in Section 2.2).

For example, features like "average value of a <lab-test>" can be generated either by Event or Numerical Table extractors, and features like "indicator of a <diagnosis>" can be generated by either Event or Categorical Table extractors. The choice of extractor depends on the format of the source data. The Time-to-Index-Date extractor is used for computing "age" at index-date, or features like "time since first <diagnosis>". The sequence of events / data table, as well as the aggregation type and observation window, are configured by parameters. An SQL query is automatically built for each of these extractors, allowing the bulk computation to be performed in the database.

On top of the above-listed "atomic" extractors, the following extractors combine and manipulate features computed by other extractors:

1. **Group extractor**: concatenates output features of multiple extractor nodes.

2. **Family extractor**: executes a given extractor across a (configurable) list of parameter configurations and concatenates the resulting output features.
3. **Filter extractor**: filters out features (e.g., sparse features). The filtering type is configurable.
4. **Transformer extractor**: transforms features (e.g., algebraic transformations or second-tier aggregation of features). The transformation type is configurable.

One of the functionalities provided by the framework is the management of train-data statistics, to support extraction from new data, e.g., during the test phase. For example, an extractor that filters sparse features may keep statistics identifying the filtered features, so the same set of features will be filtered for new data; or an extractor that computes normalized features (e.g., weight percentile). Extractors that require train-data statistics are often *context-sensitive*, that is, features' vectors computed for the same individual may change for different cohorts. Extractors that depend on a context-sensitive extractor are expected to be context-sensitive as well.

2.2. Multi-Cohort Support

Often there is an interest in analyzing two or more related cohorts; for example, all patients having a disease; women who have the disease; and women in a certain age range having the disease. The framework manages all the cohorts in a *cohort repository*, which compactly represents each cohort by a list of quartets, termed *samples*. A sample is comprised of four fields: (1) patient identifier; (2) start-date; (3) index-date; and (4) end-date. The start-date and end-date bound and limit the overall time period for which data is analyzed for the patient. The index-date partitions this time period into two: (1) *baseline* period [start-date, index-date], and (2) the *follow-up* period [index-date, end-date]. In a prediction task, features are extracted from the baseline period, while the predicted outcome is extracted from the follow-up period. The framework allows for the same individual to appear more than once in a cohort, under the restriction that the corresponding samples have different index-dates.

2.3. Speeding Up and Tracking Feature Extraction

Data analysis, and feature engineering in particular, are typically an iterative process, where different iterations may involve overlapping feature sets and/or cohorts. Our framework provides an efficient modular memoization mechanism for the extraction process, which identifies and reuses previously computed features. The extraction engine automatically assigns each extractor node with an ID that uniquely maps to the node's configuration, cohort, and dependencies in the graph. This ID is then used for associating each extractor with its cached output features and train-data statistics. Multi-cohort analysis, including cross validation and bootstrapping, often involve sub-cohorts of a larger cohort. To optimize the extraction for sub-cohorts, the memoization of every extractor that is *not* context-sensitive is performed at the level of the larger cohort. Finally, the output features include the ID of the root extractor, which can be used for tracking the configuration of the entire extraction process, e.g., for reproducibility purposes. This ID should also be used in the test phase, so the framework can locate the train-data statistics of each extractor.

We extended the framework to support Big Data analysis using Apache Spark [2]. In this extension, the feature matrix is built in a distributed manner using Spark DataFrames, and the entire extraction process can be integrated into a Spark machine learning pipeline (MLlib [3]).

3. Framework Demonstration

We demonstrate the usefulness of our feature engineering framework on a real-world, de-identified (structured) claims dataset of 320K patients, for risk factor analysis across multiple cohorts. The considered risk was for urgent care visits during a one-year period. Previously, we effectively extracted thousands of features from this dataset for applications such as risk prediction [4] and contextual anomaly detection [5]. We evaluate the predictivity of the extracted features for the outcome and compare their importance for three chronic conditions: epilepsy, diabetes, and hypertension. We use the cohort of all patients as a reference.

The baseline period was defined as the first two years of the data (2002-2004) in all four cohorts. Each of the three chronic condition cohorts was defined by requiring that patients have at least one relevant diagnosis (i.e., hypertension, diabetes, or epilepsy) during the baseline period. The outcome was defined as having at least one urgent care visit during the one-year follow-up period (2004-2005). We extracted features from all the entities available in the data: diagnosis and procedure codes, pharmacy prescriptions, lab values, and patient demographics. We used HCUP-US Clinical Classifications Software (CCS) for grouping related ICD9 (diagnoses) and CPT (procedures) codes. We utilized the National Drug Code (NDC) Directory to cross-index each NDC to its corresponding generic drug components. Sparse features with less than 100 non-zero/valid values were filtered out. We imputed missing values in lab test features using the average of existing values and used Random Forest and its Gini-importance [6] for ranking the importance of features. The results of the analysis are summarized in Table 1. The extracted features were found to be informative in every cohort (out-of-bag AUC between 0.69 and 0.72). Inspection of the top-10 ranked features in each cohort revealed that risk factors for urgent care are dominated by utilization features, such as number of claims and number of prior urgent care visit. Risk factors not directly related to utilization included “Age” and features associated with the cohort (e.g., “Epilepsy; convulsion: count” in the epilepsy cohort).

4. Conclusion

We described a framework that enhances the process of feature engineering and extraction in a multi-cohort analysis of longitudinal structured data. Our framework offers: (1) efficient management of cohorts; (2) an extendible library of reusable feature extractors; (3) a modular memoization mechanism for accelerating the extraction of overlapping feature sets; and (4) a tracking mechanism for reproducibility of former extractions. The usability of this framework was successfully used in a variety of applications, including risk analysis [4] and detection of unexpected response to treatment [5].

EHRs contain longitudinal data on patients of various types: diagnoses, drug prescriptions, lab test results, medical procedures, and more. The dynamic and irregular

nature of EHR data often requires engineering features that provide different summary statistics per patient, such as: the average value for each lab test result; or the proportion of days covered for each drug. Our extractors library enables simplified definition and reliable computation of such features for configurable time windows and cohorts. The repertoire of supported summary statistics is relatively large and is constantly growing.

One of the unique traits of the framework is its support of context-sensitive features. The framework provides infrastructure for managing train-data statistics for context-sensitive feature extractors. Moreover, its memoization mechanism, which optimizes extraction of sub-cohorts, automatically identifies context-sensitive extractors and properly handles their memoization.

We further supplemented our framework with a plug-in for Apache Spark [2] that extends our capabilities to analyze Big Data. Future work includes improving user experience via a graphical user interface that will further simplify the use of the framework, and an integrated visualization layer previously described in [5] that will allow inspection of defined cohorts as well as present longitudinal data of individual patients. Following all of the above, our framework offers an unprecedented, powerful means for analyzing and manipulating EHR data.

Table 1. Analysis results. #Positive = number of patients with positive outcome; TOP-10 = Non-utilization risk factors among top 10 ranked, excluding age

	Data statistics			Prediction model	
Cohort	Size	#Positive	#Features	AUC	TOP-10
All	320K	22.5K (7%)	949	0.69	1) Other upper respiratory infections: count
Hypertension	51.5K	5K (10%)	729	0.69	1)Essential hypertension: count 2)TRIG lab: average 3)CHOL lab: average
Diabetes	26K	2.5K (10%)	603	0.7	1)Diabetes mellitus w/o complication: count 2)TRIG lab: average 3)CHOL lab: average
Epilepsy	2K	0.27K (13%)	175	0.72	1)Epilepsy; convulsion: count 2) CHOL lab: average

References

[1] T. Tran, W. Luo, D. Phung, S. Gupta, S. Rana, R.L. Kennedy, A. Larkins, S. Venkatesh, A framework for feature extraction from hospital medical data with applications in risk prediction, BMC Bioinformatics. 15 (2014) 425. doi:10.1186/s12859-014-0425-8.

[2] Apache Software Foundation, Apache Spark, (2016). http://spark.apache.org/.

[3] X. Meng, J. Bradley, B. Yavuz, E. Sparks, S. Venkataraman, D. Liu, J. Freeman, D.B. Tsai, M. Amde, S. Owen, others, MLlib: Machine Learning in Apache Spark, J. Mach. Learn. Res. 17 (2016) 1–7.

[4] H. Neuvirth, M. Ozery-Flato, J. Hu, J. Laserson, M.S. Kohn, S. Ebadollahi, M. Rosen-Zvi, Toward Personalized Care Management of Patients at Risk: The Diabetes Case Study, in: Proc. 17th ACM SIGKDD Int. Conf. Knowl. Discov. Data Min., ACM, New York, NY, USA, 2011: pp. 395–403. doi:10.1145/2020408.2020472.

[5] M. Ozery-Flato, L. Ein-Dor, N. Parush-Shear-Yashuv, R. Aharonov, H. Neuvirth, M.S. Kohn, J. Hu, Identifying and Investigating Unexpected Response to Treatment: A Diabetes Case Study, Big Data. 4 (2016) 148–159.

[6] L. Breiman, Random Forests, Mach. Learn. 45 (n.d.) 5–32. doi:10.1023/A:1010933404324.

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-186

Development and Evaluation of a Case-Based Retrieval Service

Emilie PASCHE [a,b,1], Marcello CHINALI [c], Julien GOBEILL [a,b], Patrick RUCH[a,b]
[a] *BiTeM Group, Information Science Department, University of Applied Sciences of Western Switzerland (HES-SO, HEG), Switzerland*
[b] *SIB Text Mining, Swiss Institute of Bioinformatics, Switzerland*
[c] *Department of Pediatric Cardiology and Cardiac Surgery, "Bambino Gesù" Children Hospital, Italy*

Abstract. Identifying similar patients might greatly facilitate the treatment of a given patient, enabling to observe the response and outcome to a particular treatment. Case-based retrieval services dealing with natural language processing are of major importance to deal with the significant amount of unstructured clinical data. In this paper, we present the development and evaluation of a case-based retrieval (CBR) service tested on a collection of Italian pediatric cardiology cases. Cases are indexed and a search engine is proposed. Search functionalities, such as interactive MeSH normalization and relevance feedback, are proposed. While the qualitative evaluation aims to provide feedback and recommendations, the quantitative evaluation enables to estimate the precision of the system. In more than half of the cases and for up to two thirds of them, the system is able to suggest a similar episode of care at first rank. With an improvement of the feedback relevance strategy, we can expect an improvement of the precision. The CBR can be expanded to multilingual EHR and other fields.

Keywords. Natural language processing, ontology, electronic health record

1. Introduction

Physicians, who are facing complex diseases, show a great interest in finding populations of patients similar to their patients. Thus, they can observe the response of a particular treatment and learn about the outcomes at different points in time in a given clinical pathway. However, a substantial part of the clinical information is stored in unstructured textual contents. Therefore, tools are essential to enable the retrieval of similar cases. While case-based retrieval (CBR) tools based on structured data are numerous, less systems are managing unstructured data. Miotto *et al.* [1] describe a CBR system aiming to identify eligible patients for clinical trials. This system is based on structured data (i.e. diagnosis, medications and laboratory results) and unstructured data (i.e. clinical notes). Hsu *et al.* [2] present a CBR system dedicated to patients with intracranial aneurysm. This system uses various modalities, such as free-text clinical reports and structured data. A model-driven visualization enables to facilitate the understanding of the output.

[1] Corresponding author, Emilie Pasche, Haute Ecole de Gestion, Campus Battelle Bâtiment B, Rue de la Tambourine 17, 1227 Carouge, Switzerland; E-mail: emilie.pasche@hesge.ch.

Mourão *et al.* [3] report on a CBR system based on multimodal data. The system uses images to enrich the queries.

As part of the MD-PAEDIGREE project, a case-based retrieval (CBR) service has been developed. This service aims to find similar episodes of care based on different modalities: unstructured data (i.e. discharge summaries) and structured data (i.e. gender and age). In this paper, we present the development and evaluation of this CBR service, on a test collection of Italian pediatric patients with cardiac diseases.

2. Method

2.1. Data

The CBR is based on a set of 47,433 episodes of care, corresponding to 33,674 distinct patients consulting for cardiac pathologies. The source data originate from two Italian hospitals. Extracted episodes of care contain a discharge summary, called *clinical synthesis* in the following. Textual contents are in Italian. Demographic data (i.e. gender and age) are also retrieved.

2.2. Development of the Case-based Search Engine

The system harvests electronic health records (EHR) from the MD-PAEDIGREE infostructure with a secured API developed by GNúBILA. Normalized descriptors (i.e. MeSH) are automatically assigned to each case using MHIta [4], a service developed by HES-SO to normalize clinical texts written in Italian with MeSH descriptors. Selection of MeSH descriptors is based on a dynamic threshold strategy. Cases are then indexed using Apache Solr. At query time, MeSH descriptors are also interactively being assigned to the query. The Solr retrieval engine outputs similar cases in EHR. A weight of 0.001 is attributed to the age and gender, while the MeSH descriptors and unstructured text receive a weight of 1. The user can then assess retrieved episodes of care as relevant or not relevant. These judgements are used to reformulate the query with additional keywords based on a Rocchio algorithm [5]. Therefore, the user can obtain refined results.

2.3. Quantitative and Qualitative Evaluation

While two out of five electronic information systems are abandoned or do not respect the requirements [6], the testing and validation of a service is of major importance. Three dimensions are commonly considered: the usefulness of a medical system, its robustness and its facility of use.

The qualitative evaluation is based on the usability testing methodology [7]. The system is tested by an end-user (i.e. a MD specialized in pediatric cardiology), who performs tasks (i.e. to search for similar cases to a given case). During the whole process, the end-user is asked to verbalize his thoughts. An evaluator (i.e. a researcher) is observing and recording his comments. The data are then coded and classified by the evaluator and recommendations to improve the systems are proposed.

The quantitative evaluation is based on a benchmark, thus following the standard practice in the information retrieval domain [8]. A set of 40 queries is created. A query corresponds to the clinical synthesis of an episode of care randomly selected among the

47,433 episodes of care of the collection. An expert in cardiology manually acquires the relevance judgments. The expert executes each of the queries on the CBR and assesses the top-10 results with one of the following categories: relevant (i.e. similar to the input case) or irrelevant (i.e. judged as not similar to the query). Because this task is precision-oriented (i.e. the CBR does not aim at retrieving all the similar cases, but rather at retrieving some similar cases in order to extract useful information), we focus on precision metrics. Precision is the proportion of retrieved instances that are correct.

3. Results

3.1. GUI

The service can be accessed through the MD-PAEDIGREE portal but is restricted to allowed users due to the use of confidential data. The CBR service is a 5-step process. First, the user describes a patient with natural language, and can optionally add the age and gender of the patient. Second, the query can be interactively refined with additional keywords automatically suggested: MeSH concepts or keywords obtained by relevance feedback (only available after a first iteration). Third, the user can filter the results based on the structured data (e.g. show only boys from 3 to 10 year-old). Fourth, the similar episodes of care are displayed, ranked by relevance. To facilitate the processing by the physician, following information is displayed: demographic information (i.e. gender and age), MeSH terms automatically attributed to the clinical synthesis, clinical synthesis, a relevance score, link to the full patient history. In addition, a radio button is proposed, representing the relevance judgement. The user can then iterates to refine his query and thus obtain more relevant results, or he can expand his query to external resources (e.g. literature).

3.2. Qualitative Evaluation

The end-user appreciated the simplicity of use of the CBR service. Nevertheless, a few technical problems have arisen during the evaluation. The two main problems observed were truncated reports and the failure to answer to some queries. Those technical problems were fixed right after the evaluation session.

The automatic MeSH normalisation triggered a strong interest from the evaluator, which is familiar with the terminological resources as it is used by the MEDLINE digital library – the legacy reference for healthcare literature.

The Rocchio relevance feedback feature showed some limitations during the evaluation session. The suggested terms were reported as too general (i.e. common Italian words) or not clinically relevant. However, data analysis showed that for more than 90% of the queries, a few terms were selected.

Regarding the similar episodes of care suggested by the CBR, it was reported that the system was very efficient to retrieve similar cases when the input case was a regular case. However, the system showed difficulties to deal with the detection of the grade (e.g. normal, minor, severe, etc.), with the detection of negation or nuance (e.g. may be, unlikely, etc.), or with long and complex queries.

The evaluators also tested the preliminary version of the Rocchio-based relevance service. The results were very diverse: for a few queries, some additional relevant

documents were retrieved, for others irrelevant documents were added, while for some queries, the additional keywords did not bring any change in the ranking of the cases.

3.3. Quantitative Evaluation

Among the 40 queries, two queries were excluded due to technical failure. Eight queries returned no similar case among the top-10. Table 1 shows different measures of precision, for all queries, and for queries with at least a relevant identified answer. In more than half of the cases and for up to two thirds of them, the system is able to suggest a similar episode of care at first rank. Further, Table 2 presents the results obtained with the relevance feedback algorithm. We observe a slight improvement of the P5 and P10 with the Rocchio-based results.

Table 1. Evaluation of the first round of results returned by the CBR

	All queries (38)	**Queries with at least a relevant case (30)**
P0	0.5	0.63
P5	0.44	0.55
P10	0.42	0.54

Table 2. Evaluation of the Rocchio-based results returned by the CBR

	All queries (24)	**Queries with at least a relevant case (19)**
P0	0.5	0.63
P5	0.52	0.65
P10	0.45	0.56

4. Discussion

For eight queries, no similar case was found in the top-10. There are two hypotheses that might be considered to explain such phenomena: the system was not able to find relevant documents for these queries; the collection did not contain any relevant documents for these queries, meaning the case is so rare that there is no similar case. It is useful to highlight in this respect, that being the experiment conducted among patients with rare diseases (pediatric congenital cardiac malformations), unique cardiac phenotypes are often encountered in clinical practice, possibly explaining the lack of similarity match. If this second explanation is valid then such queries are artificially decreasing the precision of the search engine. The real precision of the system is therefore located between these lower (i.e. including the eight queries with no relevant document identified) and upper boundaries (i.e. excluding the eight queries).

The relevance feedback functionality is worth being further explored. Indeed, despite its very basic tuning at the moment of the evaluation, the quantitative evaluation showed a small positive impact on the precision. Several options are envisaged: 1) filtering of the terms suggested by the Rocchio algorithm to clinical terms only; 2) investigating negative feedback; 3) filtering words with a high document frequency using IDF (Inverse Document Frequency).

While this approach has been tested on a collection of documents in Italian, it can be expended to other languages. Indeed, the cases are automatically normalized with a terminology available in multiple languages: the MeSH terminology. Developments are being made to integrate episodes of care in English.

A limitation of our quantitative evaluation study is first the limited number of results assessed for each query and second the single expert who evaluated the results. A higher number of results assessed would enable the possibility to tune the system (i.e. to try to maximize the number of relevant results in the first positions). Indeed, as we do not know if results from position 11 are relevant, any of these results pushed in a top position after tuning would decrease the precision.

5. Conclusion

We have thus developed a case-based retrieval service dealing with several modalities (e.g. structured data, unstructured data, ontologies) and proposing various functionalities to search for similar cases (e.g. search in EHR, search in literature, relevance feedback, etc.). A methodology to develop and monitor the progress of the CBR prototype has been implemented and tested. The feedback obtained from the qualitative evaluation, despite the known rarity of the group of diseases analysed, was sufficient to improve the application regarding usability. From a quantitative point of view, the current results are already regarded as fair to support a case-based retrieval application, although several components, such as the relevance feedback service, needs fine-tuning to convince the end-users.

Acknowledgements

This experiment has been supported by the MD-PAEDIGREE project, partially funded by the European Union under the Information Communication Technologies Programme (contract number 600932). We would like also to acknowledge David Manset, Sébastien Gaspard and Nicolas Mugnier for their help with the extraction of EHRs and integration of the CBR within the MD-PAEDIGREE portal.

References

[1] R. Miotto, C. Weng, Case-based reasoning using electronic health records efficiently identifies eligible patients for clinical trials, *J Am Med Inform Assoc* **22** (2015), e141-50.
[2] W. Hsu, R.K. Taira, F. Viñuela, A.A. Bui, A Case-based Retrieval System using Natural Language Processing and Population-based Visualization, *Proc IEEE Int Conf Healthc Inform Imaging Syst Biol* **2011** (2011), 221-228.
[3] A. Mourão, E. Martins, J. Magalhães, Multimodal medical information retrieval with unsupervised rank fusion, *Comput Med Imaging Graph* **39** (2015), 35-45.
[4] P. Ruch, Automatic assignment of biomedical categories: toward a generic approach, *Bioinformatics* **22(6)** (2006), 658-64.
[5] P. Ruch, I Tbahriti, J. Gobeill, A.R. Aronson, Argumentative feedback: A linguistically-motivated term expansion for information retrieval, *Proceedings of the COLING/ACL* (2006):675-82.
[6] J. Horsky, K. McColgan, J.E. Pang, A.J. Melnikas, J.A. Linder, J.L. Schnipper, B. Middleton, Complementary methods of system usability evaluation: surveys and observations during software design and development cycles, *J Biomed Inform* **43(5)** (2010):782-90.
[7] A. Kushniruk, Evaluation in the design of health information systems: application of approaches emerging from usability engineering, *Comput Biol Med* **32(3)** (2002):141-49.
[8] S.E. Robertson, S. Walker, M. Beaulieu, Experimentation as a way of life: Okapi at TREC, *Information Processing & Management* **36(1)** (2000):95-108.

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-191

Learning Differentially Expressed Gene Pairs in Microarray Data

Xiao-Lei XIA[a], Sinead BROPHY[b], and Shang-Ming ZHOU[b,1]
[a] *School of Mechanical and Electrical Engineering, Jiaxing University, Jiaxing, P. R. China, 314001*
[b] *Farr Institute of Health Informatics Research, Swansea University Medical School, Swansea, SA2 8PP, UK*

Abstract. To identify differentially expressed genes (DEGs) in analysis of microarray data, a majority of existing filter methods rank gene individually. Such a paradigm could overlook the genes with trivial individual discriminant powers but significant powers of discrimination in their combinations. This paper proposed an impurity metric in which the number of split intervals for each feature is considered as a parameter to be optimized for gaining maximal discrimination. The proposed method was first evaluated by applying to a synthesized noisy rectangular grid dataset, in which the significant feature pair which forms a rectangular grid pattern was successfully recognized. Furthermore, applying to the identification of DEGs on colon microarray data, the proposed method demonstrated that it could become an alternative to Fisher's test for the prescreening of genes which led to better performance of the SVM-RFE method.

Keywords. Differentially expressed genes, Microarray data, Gene interactions, Machine learning

1. Introduction

In the analysis of microarray data, one of the most important tasks is the identification of differentially expressed genes (DEGs). Due to the large number of genes, univariate ranking methods have been widely employed, which can be divided into two categories: parametric approaches and model-free ones. The former category, epitomized by the *t*-test and ANOVA, assumes an underlying distribution that the samples are drawn from. Model free methods, in contrast, circumvent the assumption about data generation. To detect the DEGs, some methods have used a subset of the training sets, in which permutation of samples was implemented in order to prevent the false positive error from accumulating due to multiple testing. These methods include approaches bounding the "Family-Wise Error Rate" (FWER) which is the overall chance of one or more false positives and those controlling the "False Discovery Rate" (FDR) which is the expected percentage of false positives among the genes deemed as differentially expressed [10]. To identify important genes, some researchers have proposed gene selection methods, such as the gene pair selection approach [3], correlation-based approaches [11], Markov blanket filtering [6], minimum redundancy maximum

[1] Corresponding author, Farr Institute of Health Informatics Research, Swansea University Medical School, Swansea, SA2 8PP, UK; E-mail:s.zhou@swansea.ac.uk.

relevance [5] and uncorrelated shrunken centroid [12]. The rationale behind this scheme is that a good feature subset is highly correlated with the class and uncorrelated with each other[11]. The above algorithms as filter methods perform feature selection independently of a classifier. In contrast, wrapper feature selection methods use a classifier to evaluate a feature subset from which a classifier is trained. A number of heuristic search strategies are thus proposed, such as, estimation of distribution, sequential search, genetic algorithms [9], as well as a incremental augmenting search scheme preceded by univariate gene ranking. In gene selection, embedded methods, on the other hand, use the intrinsic property of a specific classifier to evaluate feature subsets, such as random forest induced approaches [4] and algorithms measuring the importance of genes by weight vectors respectively yielded by Support Vector Machines (SVM), known as the SVM-Recursive Feature Elimination (RFE) algorithm [7] and logistic regression [8].

However, the majority of current methods for the detection of DEGs tend to rank genes individually. The problem is that such a paradigm is very likely to dismiss the genes whose discriminant powers are trivial individually but significant jointly, as exemplified by the rectangular grid dataset contaminated by with different noise levels. Thus, in this paper we proposed an impurity metric which can efficiently identify gene pairs with good generalization performances. The novelty of this metric is that the values of each feature are split into intervals while the number of split intervals is considered as a parameter to be optimized for gaining maximal discrimination. The significance of a feature pair is measured by the sum of correctly classified training samples across all the grids on the plane. The more significant a feature pair is, the larger the sum is. The method was compared with the traditional Fisher's ratio test in the contexts of identification of DEGs. Experiments on the colon dataset [1] demonstrated that our proposed method could become an alternative to Fisher's test for the prescreening of genes which led to better performance of the SVM-RFE method.

2. Methods

Our method starts with splitting each feature into multiple intervals. Then, a pair of features divide the data into a specific number of rectangular grids. For the training samples, these grids contain either the samples from multiple classes, or the samples from a sole class, or no samples. Assuming the total number of training samples within a grid is n, among which m samples are from the positive class. The number of correctly classified samples is the maximum of m and $(n - m)$. The significant feature pair is defined as the pair that optimise the problem Eq. (1):

$$\max_{(i,j)} \sum_{k=1}^{(\#vi)^2} \Delta(i,j,k) \tag{1}$$

where i and j are the indexes of the feature pair. Representing the number of value interval for each feature by $\#vi$, $\Delta(i, j, k)$ is the number of correctly-classified samples of the k-th grid among the $(\#vi)^2$ grids that features i and j divides the data into. So the significance of a feature pair is measured by the sum of correctly-classified training samples across all the grids. The more significant a feature pair is, the larger the sum is. The least significant a feature is, the closer the sum is to half of the total number of training samples. The rationale behind the algorithm lies in our novel perspective on

the binary classification process, with which the input space is partitioned into a specific number of disjoint grids. Each grid carries a specific label which indicates the class of all, or the majority, of the inclusive training samples.

Grid search is employed to find the optimal settings of the parameters. In model selection, each classifier is trained with the selected hyper-parameters, in which the model performance is evaluated on validation data. The grid search selects the settings of the hyper-parameters that achieved the highest score in the validation procedure.

The proposed metric is similar, in terms of its methodology, to the well-known metric of gini impurity. But gini impurity is developed in the framework of decision trees and thus the range of each feature is split into two intervals, while our method allows for multiple value intervals for each feature.

3. Results of Prescreening Genes for Microarray Data

The proposed method was employed as a filter method to prescreen genes in microarray data for the identification of significant feature pair, in which the SVM with selected genes was used to classify the samples. The performance of our proposed method was compared with that of Fisher's ratio test in the SVM-RFE in terms of the prediction accuracy of the SVM established upon the selected genes. In our method, each feature pair was assigned a score as the total number of correctly-classified training samples. Assuming n genes for a microarray dataset which results in $n(n-1)/2$ pairing scores, five alternative ranking strategies were proposed:

Strategy 1: *A gene's rank is decided by the mean of the* $(n-1)$ *scores from pairing the gene with the remaining* $(n-1)$ *gene respectively.*

Strategy 2: *Among all the* $n(n-1)/2$ *pairs, find the genes pairs whose scores are among the highest. The union of these gene pairs is used as the candidate gene set.*

Strategy 3: *Select the genes pairs with the highest scores. Then, with a gene being included in one particular gene pair, the gene pairs with lower scores will be removed. The union of the resultant set of gene pairs is taken as the candidate set.*

Strategy 4: *Among all the* $n(n-1)/2$ *pairs, find the genes pairs with the lowest scores. The union of these gene pairs is excluded from further analysis.*

Strategy 5: *Select the genes pairs with the lowest scores. Then, with a gene being included in one particular gene pair, the gene pairs with higher scores will be kept. The union of the resultant set of gene pairs is excluded from further analysis.*

For microarray data with normal features less than 100 samples, it is highly recommended to employ the bootstrap resampling technique with replacement for an unbiased estimate [2]. The overall classification performance is the average of the performances on the resampled sets.

In this study, the colon dataset contains the expression values of 2000 genes with highest minimal intensity from 62 tissues. The identity of the 62 tissues is given in file tissues. There are 22 normal tissues and 40 cancerous tissues [1]. The data was subjected

to base 10 log transformation, followed by each scaled to the value range of $[-1, 1]$. The five strategies and Fisher's ratio were respectively applied to the data to reduce the number of candidate genes to 1000. The three subplots for Figure 1 from left to right corresponds to the setting of $\#vi$ at 2, 3 and 4 respectively for our proposed method. In each subplot, the six different filtering methods were highlighted in different colors.

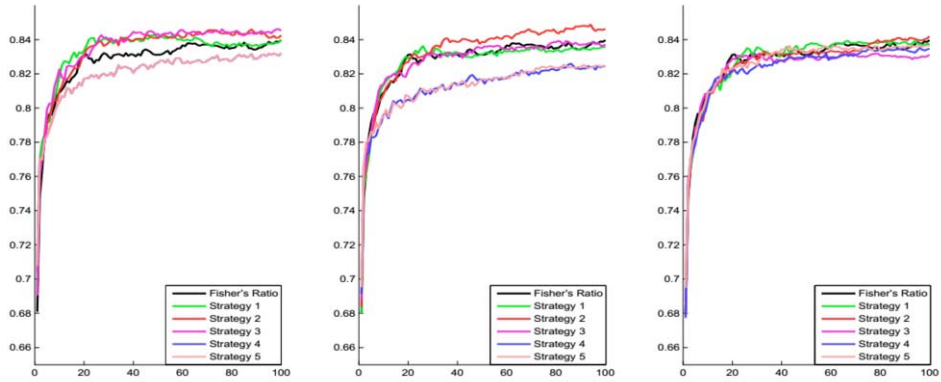


Figure 1. The performance of the SVM as a function of the number of DEGs with the regularization parameter $C = 10$: The x-axis is the number of DEGs and the y-axis represents the accuracy of the SVM classification.

Table 1 gives the top 48 pair of DEGs selected by Strategy 2 which corresponds to the red solid line in the leftmost subplot of Figure 1. In Table 1, the 1st column corresponds to the pairs with the top 12 ranking and the last three columns list a dozen of pairs respectively rank 13-th∼24-th, 25-th∼36-th,37-th∼48-th.

Table 1. The top 48 pairs of DEGs selected by Strategy 2

1st dozen	2nd dozen	3rd dozen	4th dozen
⟨Hsa.692, Hsa.549⟩	⟨Hsa.1972, Hsa.2645⟩	⟨Hsa.467, Hsa.562⟩	⟨Hsa.6080, Hsa.957⟩
⟨Hsa.823, Hsa.37937⟩	⟨Hsa.541, Hsa.692⟩	⟨Hsa.878, Hsa.8125⟩	⟨Hsa.4689, Hsa.652⟩
⟨Hsa.8147, Hsa.698⟩	⟨Hsa.678, Hsa.9235⟩	⟨Hsa.7877, Hsa.8223⟩	⟨Hsa.5398, Hsa.459⟩
⟨Hsa.831, Hsa.608⟩	⟨Hsa.832, Hsa.7728⟩	⟨Hsa.36694, Hsa.21562⟩	⟨Hsa.6039, Hsa.5392⟩
⟨Hsa.853, Hsa.8374⟩	⟨Hsa.580, Hsa.6472⟩	⟨Hsa.7498, Hsa.579⟩	⟨Hsa.2361, Hsa.6288⟩
⟨Hsa.688, Hsa.462⟩	⟨Hsa.2950, Hsa.36689⟩	⟨Hsa.951, Hsa.41208⟩	⟨Hsa.33965, Hsa.2610⟩
⟨Hsa.442, Hsa.662⟩	⟨Hsa.5971, Hsa.2715⟩	⟨Hsa.81, Hsa.421⟩	⟨Hsa.4252, Hsa.1454⟩
⟨Hsa.451, Hsa.692⟩	⟨Hsa.9972, Hsa.2097⟩	⟨Hsa.8052, Hsa.6814⟩	⟨Hsa.41282, Hsa.2996⟩
⟨Hsa.2357, Hsa.2928⟩	⟨Hsa.61, Hsa.45658⟩	⟨Hsa.24944, Hsa.57⟩	⟨Hsa.24506, Hsa.1410⟩
⟨Hsa.821, Hsa.6317⟩	⟨Hsa.316, Hsa.3331⟩	⟨Hsa.773, Hsa.1254⟩	⟨Hsa.3007, Hsa.29913⟩
⟨Hsa.42186, Hsa.8214⟩	⟨Hsa.36952, Hsa.960⟩	⟨Hsa.2654, Hsa.2821⟩	⟨Hsa.3306, Hsa.612⟩
⟨Hsa.3305, Hsa.6317⟩	⟨Hsa.8175, Hsa.627⟩	⟨Hsa.2471, Hsa.404⟩	⟨Hsa.3135, Hsa.3083⟩

It can be seen from Figure 1 that, with the setting of $\#vi = 2$, strategies 1, 2 and 3 all outperformed the Fisher's ratio. And with $\#vi = 2$, Strategies 2 and 3 still remained superior to Fisher's ratio. With $\#vi = 4$, the performances of the six methods were more or less the same although strategy 1 was slightly the best.

4. Discussions

The proposed metric assessed the importance of gene pairs in terms of the sum of correctly classified training samples across all the grids. Although standard grid search suffers from the curse of dimensionality, in this study the number of split intervals for each gene does not need to be too big. By rule of thumb, for a training dataset with n samples, the feasible range of #vi could be [2, $\sqrt{n}$]. In this study, the $\Delta(i, j, k)$ increases when the number of grids goes up from 4 (corresponding to $\#vi = 2$) to 16(corresponding to $\#vi = 4$), while decreasing for the number of grids from 16 (corresponding to $\#vi = 4$) to 36 (corresponding to $\#vi = 6$). So as a rule of thumb, the optimal $\#vi$ should be either $\#vi = 4$ or $\#vi = 5$.

5. Conclusions

In order to identify the significant feature pairs, this paper proposed an impurity metric to assess the significance of a feature based on the "purity" of the samples for the resultant subproblems. The advantage of the proposed method is that the number of value intervals for a feature is treated as a parameter to be optimized. In the identification of DEGs for microarray data as a filter strategy, the proposed method has demonstrated better performance than Fisher's ratio method.

References

[1] U. Alon, N. Barkai, D. A Notterman, K. Gish, S. Ybarra, D. Mack, and A. J Levine. Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. *Proceedings of the National Academy of Sciences*, 96(12):6745–6750, 1999.

[2] C. Ambroise and G.J. McLachlan. Selection bias in gene extraction on the basis of microarray gene-expression data. *Proceedings of the National Academy of Sciences*, 99(10):6562, 2002.

[3] T. Bo and I. Jonassen. New feature subset selection procedures for classification of expression profiles. *Genome Biology*, 3(4):0017, 2002.

[4] R. Díaz-Uriarte and A. de Andrés. Gene selection and classification of microarray data using random forest. *BMC bioinformatics*, 7(1):3, 2006.

[5] C. Ding and H. Peng. Minimum redundancy feature selection from microarray gene expression data. Journal of Bioinformatics and Computational Biology, 3(2):185–206, 2005.

[6] O. Gevaert, F.D. Smet, D. Timmerman, Y. Moreau, and B.D. Moor. Predicting the prognosis of breast cancer by integrating clinical and microarray data with Bayesian networks. *Bioinformatics*, 22(14), 2006.

[7] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik. Gene selection for cancer classification using support vector machines. *Machine Learning*, 46(1):389–422, 2002.

[8] S. Ma and J. Huang. Regularized roc method for disease classification and biomarker selection with microarray data. *Bioinformatics*, 21(24):4356–4362, 2005.

[9] C.H. Ooi and P. Tan. Genetic algorithms applied to multi-class prediction for the analysis of gene expression data. *Bioinformatics*, 19(1):37–44, 2003.

[10] V.G. Tusher, R. Tibshirani, and G. Chu. Significance analysis of microarrays applied to the ionizing radiation response. *Proceedings of the National Academy of Sciences*, 98(9):5116–5121, 2001.

[11] E.J. Yeoh, M.E. Ross, S.A. Shurtleff, W.K. Williams, D. Patel, R. Mahfouz, F.G. Behm, S.C. Raimondi, M.V. Relling, A. Patel, et al. Classification, subtype discovery, and prediction of outcome in pediatric acute lymphoblastic leukemia by gene expression profiling. *Cancer Cell*, 1(2):133–143, 2002.

[12] K. Yeung and R. Bumgarner. Multiclass classification of microarray data with repeated measurements: application to cancer. *Genome Biology*, 4(12):R83, 2003.

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-196

Developing a Manually Annotated Corpus of Clinical Letters for Breast Cancer Patients on Routine Follow-Up

Graham PITSON[a], Patricia BANKS[b], Lawrence CAVEDON[c], Karin VERSPOOR[d,1]
[a]*Barwon Health, Geelong VIC Australia*
[b]*Peter MacCallum Cancer Centre, Melbourne VIC Australia*
[c]*RMIT University, Melbourne VIC Australia*
[d]*The University of Melbourne, Melbourne VIC Australia*

Abstract. This paper introduces the annotation schema and annotation process for a corpus of clinical letters describing the disease course and treatment of oestrogen receptor positive breast cancer patients, after completion of primary surgery and radiotherapy treatment. Concepts related to therapy, clinical signs, and recurrence, as well as relationships linking these, are identified and annotated in 200 letters. This corpus will provide the basis for development of natural language processing tools for automatic extraction of key clinical factors from such letters.

Keywords. clinical letter corpus, annotated corpus, cancer follow-up

1. Introduction

In recent years, significant progress has been made in adoption of electronic medical records (EMR) with the aim of improving information access and flow within healthcare. However, data transfer between clinicians in different facilities remains problematic. In situations of shared care, a common form of communication is via letters between physicians. Clinical letters are generally free text documents discussing major aspects of care delivered, the current status of the patient and plans for future care. Even when stored in an EMR, their unstructured format means that details of patient care remain hidden, difficult to retrieve and unavailable for automated analysis.

Cancer treatment and follow-up is a significant part of any health care system, and breast cancer is a common malignancy. After initial therapy, many patients with breast cancer may receive outpatient based endocrine therapy for several years. This may involve clinic visits to different health care providers, and medications can be prescribed from multiple sources. In this situation, follow-up care letters shared between clinicians are a major source of clinical data regarding treatment outcomes.

This study examined a corpus of clinical letters regarding breast cancer patients on routine follow-up with the aim of (1) identifying major themes and text patterns in the clinical letters and (2) developing a reference corpus for subsequent analysis of such letters using automated natural language processing methods.

[1] Corresponding author, School of Computing and Information Systems, The University of Melbourne, Melbourne VIC 3010, Australia; karin.verspoor@unimelb.edu.au

2. Background

The follow-up care of cancer patients is often shared between clinicians and health services and information swapped via clinical documents. Natural language processing (NLP) has been shown to be able to extract information from such unstructured texts [4]. NLP pipelines have been used to identify breast cancer recurrence in clinical notes combined with pathology and radiology reports [1], as well as endocrine drug therapy patterns from EMR notes [5]. In each case, limited and predefined items of interest were extracted from the clinical history, and required the development of a corpus of texts in which those items were annotated, for training and evaluation of the NLP methods [2].

Whether NLP can be used to extract meaning more broadly from clinical notes is ultimately dependent on note content. A review of the literature on communication between doctors summarised expectations of timeliness and areas of content but also noted that in general, content was often considered to be inadequate [7]. Topics thought to be important in letters between doctors discussing patients with cancer include diagnosis and stage, treatments and side effects, prognosis, clinical status and follow-up plans [6]. However, there has been little analysis of the actual content of clinical oncology letters shared between clinicians.

3. Data: Clinical Letters

The clinical cohort was a group of early stage breast cancer patients treated in a single cancer centre and seen in the medical oncology or radiotherapy clinics. 200 clinical letters were identified for inclusion in the corpus.

Document selection: The patient cohort comprised patients diagnosed with a hormone receptor positive breast cancer in 2009 or 2010, who were treated and followed up in a large regional cancer service. The cohort all had oestrogen receptor positive breast cancers, had received surgery and radiotherapy but no chemotherapy and would normally be offered endocrine therapy. Selected letters were all dated after treatments such as surgery or radiotherapy were completed, aiming to capture the period where endocrine therapy alone was used. Each patient required a minimum of 3 follow-up letters within that time period to be considered.

The documents generally consisted of a narrative summary of the clinic visit, major clinical issues and future follow-up plans. Some letters contained a table summarising disease stage and treatment, or the time course of major clinical events. Of the 200 letters, 134 were from the medical oncology clinic and 66 from the radiation oncology clinic, with 64% written by consultants and the remainder by advanced specialty trainees.

Document pre-processing and de-identification: Before annotation each document was converted from Word format into plain text for use in the brat annotation tool (`http://brat.nlplab.org/`). Letters were de-identified at this stage. De-identification included removing patient details as well as clinician names and clinic or other addresses. Although each patient had multiple letters in the corpus the aim was to annotate each letter independently and so document names were obfuscated to remove sequencing order or other identifiable patterns. Text formatting was identified and converted into tags at the time of conversion in order to preserve potential additional meaning (e.g. bold summaries or tabular data).

Letter structure: All letters contained a standard table as a header containing patient details, patient record number and date seen. Many letters contained additional formatted text. Approximately 45% of letters had a bold text header containing a summary of the diagnosis and staging, and 73% contained a table outlining a basic chronology of the major events from diagnosis to the date the letter was created.

Most letters (over 90%) were addressed to the local family physician with others sent to the treating surgeon. An additional addressee, typically to medical records or to other direct members of the treating team, was common.

4. Annotation

A schema for annotation of this set of clinical letters was developed. The schema defines major themes expected to be covered in letters for this patient population. Spans of text corresponding to specific concepts were identified and marked in the letters, and relations between them were captured where possible.

4.1. *Annotation Schema*

There were 6 groups of *Concept* annotations, as follows (detailed in Table 1):

- **Therapy related:** mentions of endocrine therapy (the expected treatment in this corpus) but also other cancer and non-cancer related treatments.
- **Clinical:** comments on clinical signs or symptoms, names of tests, test results associated with clinical findings, indications of the severity of clinical findings (or their absence) and note of comorbidities.
- **Recurrence:** statements indicating possible recurrence or statements indicating lack or absence of recurrence (expected to be a far more common event).
- **Follow-up:** plans for further appointments with the letter sender.
- **Discussion:** statements that noted some discussion had taken place around treatments. The purpose of this annotation was to potentially allow distinction between mention of other concepts (such as endocrine therapy) in the context of a discussion rather than as part of current therapy.
- **Supporting:** additional info, such as time points indicating currency of therapy (e.g., prior, current, future).

Relations were also defined, to link concepts within four main areas:

- **Therapy:** Linking therapy to timing, toxicities or interventions for recurrence.
- **Clinical:** Linking clinical findings to severity, comorbidities or negation.
- **Recurrence:** Linking possible recurrence statements to clinical findings.
- **Results-discussion:** Linking the discussion to the context of discussion.

4.2. *Annotation Process and Merging*

Two cancer physicians – a radiation oncologist and an advanced trainee in medical oncology – annotated the corpus using the schema. The first annotator (GP) created the schema and iteratively modified it through regular meetings while annotating a subset of the total corpus. The two annotators then jointly annotated a number of documents to clarify any initial misunderstandings.

Table 1: Annotation statistics for each concept type, including total number, number of documents with at least one annotation of that type (Doc Cov=coverage), the average number of annotations per covered document, and inter-annotator agreement (IAA, using F1-measure).

Concept	Count	Doc Cov (%)	Num/Doc	IAA
Endocrine therapy	390	186 (93)	2.1	0.97
Other therapy (non-cancer)	113	57 (29)	2.0	0.88
Clinical finding	211	136 (68)	1.6	0.94
Clinical severity	69	53 (27)	1.3	0.95
Name of test	305	150 (75)	2.0	0.97
Test result	114	82 (41)	1.4	0.93
Comorbidity (not breast cancer)	63	47 (24)	1.3	0.86
No evidence of recurrence	472	177 (89)	2.7	0.84
Follow-up with this physician	191	190 (95)	1.0	0.98
Discussion noted	67	41 (21)	1.6	0.80
Context/Basis of discussion	53	40 (20)	1.3	0.84
Timing of therapy	291	182 (91)	1.6	0.91
Time point	400	189 (95)	2.1	0.95

Once all documents had been annotated, reviewed and checked by each annotator, a single merged corpus was created. For overlapping annotations of the same concept, the longest text span was taken. Concepts annotated by just one annotator were added to the merged set as long as the new concept did not clash (overlap) with an existing concept. Relations were added automatically where the relation and both related concepts were to be added to the merged set.

5. Results

A summary of the annotations is presented in Table 1 (for concepts) and in Table 2 (for relations). A total of 2739 concept and 898 relation annotations were made across the 200 clinical letters in the corpus. Inter-Annotator Agreement (IAA) was assessed using the F1-measure, the harmonic mean (β=1) of Precision (or positive predictive value, $\frac{TP}{TP+FP}$) and Recall (or sensitivity, $\frac{TP}{TP+FN}$), treating one annotator's annotations as truth and evaluating the overlap of the other annotator's annotations with those. This measure has advantages over other measures in contexts such as this where the expected number of true negatives – text lacking annotations – is ill-defined [3].

Concept annotation agreement was measured using an exact match (where both the annotated text span and annotated concept agreed; results not shown) and an overlap match (where the annotated concept matched but the text spans of the annotators overlapped; shown in Table 1).

Concepts: The concepts of No recurrence, follow-up and Timing of therapy had low exact F-measures of 0.56, 0.72 and 0.70 but overlap F-measures of 0.85, 0.98 and 0.92 respectively. Most of the exact match difference was caused by inclusion of text identifying the patient (such as the de-identified patient name or pronoun such as 'she') by one of the annotators. The Discussion concept had low exact match agreement; this was due to differing interpretations of the annotation guidelines. Accounting for overlap, the Discussion concept F-measure improved to 0.80.

Table 2. Relation annotation statistics for each concept type, reported as above.

Relation	Count	Doc Cov (%)	Num/doc	IAA
Timing of therapy	314	180 (90)	1.7	0.89
Next review timing	165	164 (82)	1	0.98
Toxicity of endocrine therapy	159	118 (59)	1.4	0.97
Timing of test	191	118 (59)	1.6	0.92
Result of test	133	82 (41)	1.6	0.92
Time point of event	78	57 (29)	1.4	0.82
Severity of toxicity	74	53 (27)	1.4	0.93
Result – recurrence (or lack) link	51	49 (25)	1.0	0.7
Discussion – context link	43	32 (16)	1.5	0.72
Therapy – condition treated link	29	25 (13)	1.2	0.75
Time link	20	20 (10)	1.0	0.80

Relations: Relations were considered a match if the relation type and both linked concepts matched. The lowest F-measure is seen in relations linking concepts that themselves have relatively low annotation agreement.

6. Discussion and Conclusions

This study has shown that annotation of a corpus of clinical documents in breast cancer is feasible. Allowing for concept matching based on overlapping text spans resulted in good agreement, with F-measures of 0.80 or above. Themes are consistent across letters and include the important clinical information identified by [6]; around 90% of clinical letters of early stage breast cancer patients mention the prescribed endocrine therapy, disease status and follow-up plans. Toxicity of therapy (or absence) and tests performed are noted less frequently.

This quantitative content analysis of clinical letters in breast cancer patient follow-up is novel and has confirmed that these unstructured documents contain valuable clinical information. Future work will develop natural language processing techniques using this data to enable automatic extraction of this information.

References

[1] D. Carrell, S. Halgrim, et al. Using natural language processing to improve efficiency of manual chart abstraction in research: the case of breast cancer recurrence. Am J Epidem, 179(6):749–758, Mar 2014.
[2] S. L. DuVall et al. Creating reusable annotated corpora with the clinical document architecture. In 44th Hawaii Intl Conf System Sciences, pages 1–10, Jan 2011.
[3] G. Hripcsak and A. S. Rothschild. Agreement, the f-measure, and reliability in information retrieval. Journal of the American Medical Informatics Association, 12(3):296, 2005.
[4] A. Jha. The promise of electronic records: around the corner or down the road? JAMA, 306(8):880–881, Aug 2011.
[5] G. Savova et al. Automated discovery of drug treatment patterns for endocrine therapy of breast cancer within an electronic medical record. JAMIA, 19(e1):e83–89, Jun 2012.
[6] M. Tattersall, P. Butow, J. Brown, and J. Thompson. Improving doctors' letters. Med J Aust, 177(9):516–520, Nov 2002.
[7] P. Vermeir, D. Vandijck, et al. Communication in healthcare: a narrative review of the literature and practical recommendations. Int J Clin Pract, 69(11):1257–1267, Nov 2015.

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-201

Automated Diagnosis Coding with Combined Text Representations

Stefan BERNDORFER [a,1] and Aron HENRIKSSON [b]
[a] *Faculty of Computer Science, University of Vienna, Austria*
[b] *Department of Computer and Systems Sciences, Stockholm University, Sweden*

Abstract. Automated diagnosis coding can be provided efficiently by learning predictive models from historical data; however, discriminating between thousands of codes while allowing a variable number of codes to be assigned is extremely difficult. Here, we explore various text representations and classification models for assigning ICD-9 codes to discharge summaries in MIMIC-III. It is shown that the relative effectiveness of the investigated representations depends on the frequency of the diagnosis code under consideration and that the best performance is obtained by combining models built using different representations.

Keywords. Electronic health records, diagnosis coding, predictive modeling

1. Introduction

The digitization of healthcare brought about by the adoption of electronic health record (EHR) systems has made vast amounts of data available for processing by computers [1]. Secondary use of EHR data enables the efficient and often effective provision of clinical decision support at the point of care by building predictive models that learn from large-scale observations of historical data to, e.g., automatically assign or suggest diagnosis codes. Automating the process of diagnosis code assignment can drastically reduce healthcare costs but is challenging due to the inherent difficulty of predicting one or more labels from a large set of classes [2]: ICD-9, for instance, includes around 14,000 unique codes organized in a hierarchical fashion. As a result, many studies have limited the task in some way, e.g. by focusing on a small subset of codes, as in the CMC challenge [3], or on a specific outcome such as mortality [4] or adverse drug events [5]. In reality, the number of distinct combinations of diagnosis codes in EHRs is extremely large and the distribution of codes highly skewed, both of which present challenges for supervised learning approaches [6]. To address these, Perotte et al. [7] proposed a classification strategy that exploits the hierarchy of ICD-9, demonstrating improved performance over a flat prediction model (F_1: 0.29 vs. 0.21), while a similar approach improved performance on adverse drug event detection [8]. In this study, we continue to explore these classification strategies, while focusing on yet another key issue: text representation. In a classification setting, a document is often represented as a bag of words, i.e. using simple (weighted) frequencies. Although such shallow representations often yield competitive performance, deep representations that account for the semantics of words have been proposed, improving performance on various

[1] Corresponding author: stefan.berndorfer@gmx.net

diagnosis coding tasks [9,10]. These representations derive vector representations of words (embeddings) based on their distribution in different contexts: the assumption is that words appearing in similar contexts (i.e. co-occurring with overlapping sets of words) have similar meanings. Here, we show that the relative effectiveness of these representations is related to the frequency of the considered class and that overall performance can be improved by combining shallow and deep text representations.

2. Methods & Materials

We investigate the use of various predictive models for ICD-9 coding of discharge summaries. Two text representations – shallow and deep – are provided to the learning algorithm and their effectiveness, w.r.t. predictive performance, is analyzed for diagnosis codes of varying frequency. Several strategies for combining predictive models that exploit different representations are then explored; the entire analysis is conducted using two classification strategies previously proposed in the literature.

Text Representations: Two popular text representations are used: (1) a shallow representation describing each document as a *bag-of-words* (BoW), i.e. the (weighted) frequency distribution of words in some vocabulary, here defined as the 10,000 words with the highest *Term Frequency - Inverse Document Frequency* (TF-IDF) scores in the training data; (2) a deep representation describing each document as a TF-IDF-weighted sum of semantic vectors that have been learned using the *continuous bag-of-words* (CBOW) model of *Word2Vec* (W2V) [11]. The CBOW model trains a single-layer neural network that learns to predict words based on their contexts, i.e. adjacent words within a symmetric window of a given size; the parameters learned in the hidden layer give us semantic vector representations of words.

Combination Strategies: Once predictive models have been trained using a given representation, they can be combined in an attempt to improve performance. A distinction exists between early fusion and late fusion. In the former, the combination takes place prior to learning, typically by combining feature sets. Here, a combination strategy named *Fusion* is investigated, in which features from the two representations are simply concatenated prior to learning a single predictive model. Various late fusion strategies are also explored. *Select One* chooses a representation and the corresponding model based on the observed best performance for the diagnosis code within a certain frequency interval. Two other strategies are based on simple set operations: *Union* takes the union of the predictions, while *Intersection* takes the intersection of the predictions made by the two models. Finally, *Probability Averaging* takes a weighted average of the class probabilities produced by the models; here, the weights are determined by the observed predictive performance scores for diagnosis codes within a certain frequency interval. A fitted sigmoid was used to obtain probability estimates from the trained Support Vector Machine (SVM) models [12].

Classification Models: Two classification models are used[2] [7]: the *flat SVM model* uses all available training examples, while the *hierarchical SVM model* exploits the ICD- 9 hierarchy[3]. In both settings, the multi-label problem is binarized with a one-versus-all model per diagnosis code. In the flat classification model, documents to which a given diagnosis code has been assigned serve as positive examples and all

[2] Based on PhysioNet project: https://physionet.org/works/ICD9CodingofDischargeSummaries
[3] http://bioportal.bioontology.org/ontologies/ICD9CM

others as negative examples. In the hierarchical model, codes are augmented by their ancestors as follows. Training is carried out from the root downwards: only codes from the parent's sub-tree are considered, where all instances rooted in the code itself serve as positive examples and the remaining ones as negative examples. The prediction follows the same hierarchical procedure: if a parent node has been predicted as negative, no child can be positive, while only leaf nodes serve as final predictions [13].

Experimental Setup: The experiments were conducted using data from the Medical Information Mart for Intensive Care III (MIMIC-III) [14], a publicly available database comprising de-identified health data for over 40,000 critical care patients. All discharge summaries and assigned ICD-9 diagnosis codes were extracted from the database. Codes occurring fewer than 50 times were filtered out, resulting in 59,531 non-empty discharge summaries with at least one assigned diagnosis code. The discharge summaries were tokenized, part-of-speech tagged and lemmatized, while common stopwords[4] were removed. The preprocessed corpus has a vocabulary size of around 125,000, with approximately 44 million instances. The average length of a discharge summary is 742 words (± 435.3). There are 1,301 distinct ICD-9 codes that occur a total of 634,375 times, resulting in an average of 10.66 (± 5.74) codes per summary. The code distribution is strongly skewed towards low-frequent codes: 83% of the codes occur in less then 1% of the discharge summaries. For classification, the LibLinear SVM [15] implementation was used and all representations were L2-normalized. The dataset was divided into a training set (80%), a development set (10%) and an evaluation set (10%). The following parameters were optimized using 5-fold cross-validation on the training set: the window size and dimensionality of the W2V models, as well as the c-value of the linear SVM. To limit the parameter optimization procedure, a sequential approach was taken whereby, first, the window size (5, 10, 25, 50, 100, 150, 200, 250) was optimized using a dimensionality of 200; then, the dimensionality was successively increased[5] (200, 400, 600, 800, 1000); finally, various c-values were explored (2^x, where $x \in \{0, 1, 2, 3, 4, 5\}$). The training set was also used for comparing the BoW and W2V representations: with the hypothesis that the effectiveness of a given representation may depend on the frequency of a diagnosis code, the predictive performance was analyzed in three subsets of the training set, corresponding to tertiles based on code frequency. Models were then trained on the entire training set and, based on the observations of the tertile analysis, the combination strategies were evaluated on the development set. Finally, the best single and combination models were trained on the tuning and development sets and compared on the evaluation set, where McNemar's test [17], with one degree of freedom, was used to verify the statistical significance of the results.

3. Results

The parameter tuning favored a large window size and dimensionality for the W2V spaces and various c-values for the SVM models (Figure 1). The tertile analysis shows that the predictive performance strongly decreases with a lower code frequency (Figure 2). The shallow BoW model performs better on high-frequent codes; however, for

4 http://www.ranks.nl/stopwords [Accessed October 24, 2016]

5 Increasing the dimensionality of semantic spaces can lead to improved performance [16].

medium- and low-frequent codes, the deep W2V representation outperforms BoW in both classification models.

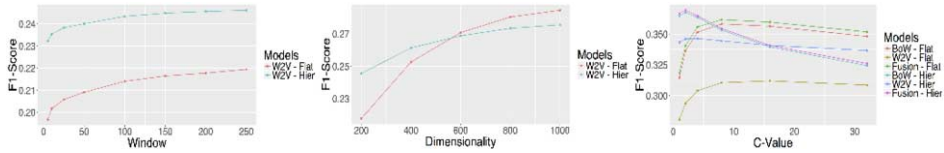


Figure 1. Parameter tuning: W2V window size, W2V dimensionality, SVM c-value.

A comparison of various combination strategies showed that the Union and Fusion models performed best in both classification models (Figure 2). The Union model leads to increased recall at the expense of precision, but outperformed the Fusion model in terms of F_1-score. When comparing the best single model with the best combined model on unseen data, the following was observed. In the flat model, BoW achieved 58.68% precision, 29.96% recall and 36.95% F_1-score. The Union model achieved 55.10% precision, 33.74% recall and 39.16% F_1-score. In the hierarchical model, the BoW representation achieved 43.96% precision, 35.98% recall and 37.97% F_1-score. The Union combination achieved 40.08% precision, 41.69% recall and 39.25% F_1-score. In both settings, the Union prediction model outperformed the BoW representation in terms of F_1-score, by 2.21 points in the flat and 1.28 points in the hierarchical setting. McNemar's test applied independently within each classification model showed that the differences in performance were significant ($p < 0.01$).

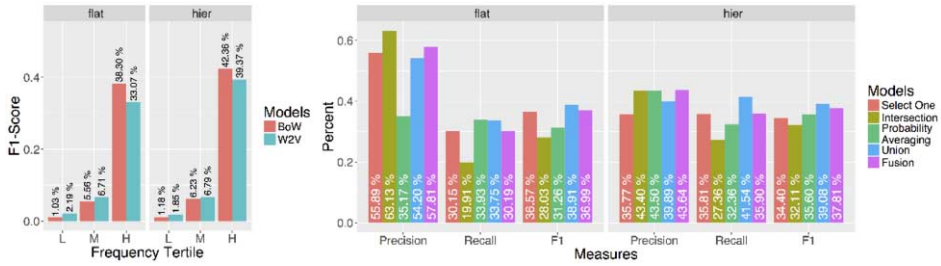


Figure 2. Results: Tertile Analysis and Combination Strategies.

4. Discussion

The tertile analysis revealed that the relative effectiveness of the two investigated text representations depends on the frequency of the diagnosis code under consideration. The deep representation outperformed the shallow counterpart for rare and medium-frequency codes, which can be explained by the lack of training examples to learn from: distributional semantics can then help the classifier to exploit similarities in word meaning between different surface tokens. The complementary nature of the representations was exploited by combining models trained using different representations. A number of combination strategies were evaluated, with the rather naive strategy of taking the union of the predictions outperforming the alternatives. This late fusion strategy hence outperformed the early fusion strategy; in another study on diagnosis code assignment where various late fusion strategies were compared to early fusion, this was not the case [18]. The combined model significantly

outperformed the best single model and the results are substantially higher than those presented in [7], in part due to properly tuning the parameters but largely as a result of combining text representations. The performance gain of the Union model can be attributed to the fact that in large clinical datasets, an increase in recall of frequent codes mostly affects performance gains in terms of F_1-score [6].

References

[1] Carol Friedman, Thomas C. Rindflesch, and Milton Corn. Natural language processing: State of the art and prospects for significant progress, a workshop sponsored by the National Library of Medicine. *Journal of Biomedical Informatics*, 46(5):765–773, October 2013.

[2] Mary H. Stanfill, Margaret Williams, Susan H Fenton, Robert A Jenders, and William R Hersh. A systematic literature review of automated clinical coding and classification systems. *Journal of the American Medical Informatics Association*, 17(6):646–651, November 2010.

[3] John P. Pestian, Christopher Brew, Pawel Matykiewicz, D. J. Hovermale, Neil Johnson, K. Bretonnel Cohen, and Wlodzislaw Duch. A shared task involving multi-label classification of clinical free text. In: Proc. of BioNLP 2007, p. 97–104. Association for Computational Linguistics, 2007.

[4] Zhengping Che, Sanjay Purushotham, Kyunghyun Cho, David Sontag, and Yan Liu. Recurrent Neural Networks for Multivariate Time Series with Missing Values. arXiv preprint arXiv:1606.01865, 2016.

[5] Jing Zhao, Aron Henriksson, Lars Asker, and Henrik Boström. Predictive modeling of structured electronic health records for adverse drug event detection. *BMC Medical Informatics and Decision Making*, 15(Suppl 4):S1, 2015.

[6] Ramakanth Kavuluru, Anthony Rios, and Yuan Lu. An empirical evaluation of supervised learning approaches in assigning diagnosis codes to electronic medical records. *Artificial Intelligence in Medicine*, 65(2):155–166, October 2015.

[7] Adler Perotte, Rimma Pivovarov, Karthik Natarajan, Nicole Weiskopf, Frank Wood, and Noémie Elhadad. Diagnosis code assignment: models and evaluation metrics. *Journal of the American Medical Informatics Association*, 21(2):231–237, March 2014.

[8] Jing Zhao, Aron Henriksson, and Henrik Boström. Cascading adverse drug event detection in electronic health records. In: International Conference on Data Science and Advanced Analytics (DSAA), 2015.

[9] Aron Henriksson, Martin Hassel, and Maria Kvist. Diagnosis code assignment support using random indexing of patient records – a qualitative feasibility study. In: Conference on Artificial Intelligence in Medicine, Springer, 2011, p. 348–352.

[10] Aron Henriksson, Jing Zhao, Henrik Boström, and Hercules Dalianis. Modeling heterogeneous clinical sequence data in semantic space for adverse drug event detection. In: International Conference on Data Science and Advanced Analytics (DSAA), 2015.

[11] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781, 2013.

[12] John C. Platt. Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods. In: Advances in Large Margin Classifiers, pages 61–74. MIT Press, 1999.

[13] Yitao Zhang. A Hierarchical Approach to Encoding Medical Concepts for Clinical Notes. In: Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics on Human Language Technologies: Student Research Workshop, HLT-SRWS '08, pages 67–72, Stroudsburg, PA, USA, 2008. Association for Computational Linguistics.

[14] Alistair E.W. Johnson, Tom J. Pollard, Lu Shen, Li-wei H. Lehman, Mengling Feng, Mohammad Ghassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi, and Roger G. Mark. MIMIC-III, a freely accessible critical care database. *Scientific Data*, 3:160035, May 2016.

[15] Rong-En Fan, Kai-Wei Chang, Cho-Jui Hsieh, Xiang-Rui Wang, and Chih-Jen Lin. LIBLINEAR: A library for large linear classification. *Journal of Machine Learning Research*, 9(Aug):1871–1874, 2008.

[16] Aron Henriksson and Martin Hassel. Optimizing the dimensionality of clinical term spaces for improved diagnosis coding support. In: Proceedings of Louhi Workshop on Health Document Text Mining and Information Analysis, 2013.

[17] Thomas G. Dietterich. Approximate Statistical Tests for Comparing Supervised Classification Learning Algorithms. 10(7):1895–1923, October 1998.

[18] Aron Henriksson, Jing Zhao, Henrik Boström, and Hercules Dalianis. Modeling electronic health records in ensembles of semantic spaces for adverse drug event detection. In: IEEE International Conference on Bioinformatics and Biomedicine (BIBM), p. 343–350, 2015.

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-206

Using Statistics and Data Mining Approaches to Analyze Male Sexual Behaviors and Use of Erectile Dysfunction Drugs Based on Large Questionnaire Data

Zhi QIAO[a, 1], Xiang LI[a], Haifeng LIU[a], Lei ZHANG[b], Junyang CAO[b], Guotong XIE[a], Nan QIN[b], Hui JIANG[c] and Haocheng LIN[c]
[a] *IBM Research - China*
[b] *Pfizer Investment Co. Ltd.- China*
[c] *Peking University Third Hospital, Beijing, China*

Abstract. The prevalence of erectile dysfunction (ED) has been extensively studied worldwide. Erectile dysfunction drugs has shown great efficacy in preventing male erectile dysfunction. In order to help doctors know drug taken preference of patients and better prescribe, it is crucial to analyze who actually take erectile dysfunction drugs and the relation between sexual behaviors and drug use. Existing clinical studies usually used descriptive statistics and regression analysis based on small volume of data. In this paper, based on big volume of data (48,630 questionnaires), we use data mining approaches besides statistics and regression analysis to comprehensively analyze the relation between male sexual behaviors and use of erectile dysfunction drugs for unravelling the characteristic of patients who take erectile dysfunction drugs. We firstly analyze the impact of multiple sexual behavior factors on whether to use the erectile dysfunction drugs. Then, we explore to mine the Decision Rules for Stratification to discover patients who are more likely to take drugs. Based on the decision rules, the patients can be partitioned into four potential groups for use of erectile dysfunction: high potential group, intermediate potential-1 group, intermediate potential-2 group and low potential group. Experimental results show 1) the sexual behavior factors, erectile hardness and time length to prepare (how long to prepares for sexual behaviors ahead of time), have bigger impacts both in correlation analysis and potential drug taking patients discovering; 2) odds ratio between patients identified as low potential and high potential was 6.098 (95% confidence interval, 5.159-7.209) with statistically significant differences in taking drug potential detected between all potential groups.

Keywords. Statistics, Data mining, Erectile dysfunction, Drug Therapy

1. Introduction

Erectile dysfunction is sexual dysfunction characterized by the inability to develop or maintain an erection of the penis during sexual activity in humans [1]. The prevalence of erectile dysfunction (ED) has been extensively studied worldwide [2,5]. Erectile dysfunction drugs, such as Sildenafil, has shown great efficacy in preventing male erectile dysfunction. However, in China, the use of erectile dysfunction drugs greatly

[1] Corresponding author. E-mail: qzbj@cn.ibm.com.

depends on patients' consciousness. Therefore, it is crucial to analyze who actually take erectile dysfunction drugs and the relation between sexual behaviors and drug use [4, 3]. Existing clinical studies usually used descriptive statistics and regression analysis based on small volume of questionnaire data. However, the studies did not clearly unravel the characteristic of patients who take erectile dysfunction drugs.

In this paper, we use statistics, regression analysis and data mining approaches to analyze the relation between male sexual behaviors and use of erectile dysfunction drugs based on the 48,630 sexual behavior questionnaires. We firstly analyze the impact of multiple sexual behavior factors on whether to use the erectile dysfunction drugs. Then, we explore to mine the Decision Rules for Stratification to discover patients who are more likely to take drugs. In order to learn the impact of sexual behavior factors, we firstly use the Pearson correlation coefficient to compute the univariate correlation between factors and outcome. Furthermore, we use multivariate logistic regression to measure the impact of factors on the outcome. Then, we use Classification And Regression Tree (CART) to learn the decision rules to discover patients who are more likely to take drugs. The learnt knowledge can help doctors know drug taken preference of patients and better prescribe. The insight discovered from the work is validated by clinical professionals and is deemed to benefit the related deeper clinical research.

2. Material & Method

By web-based digital questionnaires, data were collected in China. The participants were the web users accessing the digital questionnaire by Internet. No personal identifying information was obtained from any subject. The questionnaire requested information on the frequency, erection hardness, time length to prepare (how long to prepares for sexual behaviors ahead of time), and whether to use erectile dysfunction drugs and so on. Overall, 154,003 questionnaires were collected where 48,630 of them completed the questions in questionnaires. For decision rules learning, the 39,032 questionnaire records (80% random sampling from full data) are used to develop the model (training cohort) and the remain 9,598 questionnaire records (20% random sampling from full data) are used to test the model (testing cohort).

Based on the large questionnaire data, we analyze impact of multiple sexual behavior factors for outcome and mine the Decision Rules for Stratification to discover patients who are more likely to take drugs.

2.1. Factor Impact Analysis

We firstly use the Pearson correlation coefficient [7] to measure the correlation between the outcome (whether to use the erectile dysfunction drugs) and each factor. The correlation coefficient is a value that quantifies the dependence between factor and outcome.

In order to further associate the outcome with several factors and examine the potential confounding effect of certain factor, the logistic regression is usually applied. In the logistic regression learning, the derived odds ratio is similar to, but somewhat different from, the more commonly understood relative risk or risk ratio.

2.2. Decision Rules Learning

The CART method is an empirical, statistical technique based on recursive partitioning analysis [6]. Unlike multivariable logistic regression, it is well suited to the generation of clinical decision rules. Furthermore, because it does not require parametric assumptions, it can handle numerical data that are highly skewed or multimodal and categorical predictors with either an ordinal or nonordinal structure. The CART method involves the segregation of different values of classification variables through a decision tree composed of progressive binary splits. Every value of each predictor variable is considered as a potential split, and the optimal split is selected based on impurity criterion (the reduction in the residual sum of squares due to a binary split of the data at that tree node). This process continues with both tree building and pruning until statistical analysis indicates that the tree fits without overfitting the information contained in the data set. As a result, CART analysis produces decision trees that are simple to interpret and may be applied at the bedside.

The ability of the derived decision tree is to determine users to have low, intermediate-1, intermediate-2, and high potential for drug taking. The users from the test cohort were classified into potential groups based on the learnt CART tree. For the outcome of taking drugs, the ORs and 95% CIs between potential groups were determined (logistic regression in SPSS version 24, IBM Institute Inc).

3. Results and Discussion

Four factors are obtained from the questionnaire data, Age, Frequency of sexual activities, Erectile hardness and Time length to prepare (how long to prepares for sexual behaviors ahead of time). Because of the question request style, all factors are ordinal values. Such as, the values of Age are <20 years old, 20~30, 30~40, 40~50, 50~60 and >60; the values of Frequency (Monthly) are 0~2 times, 3~4, 5~6, 7~8, 8~10 and >10; the values of Erectile hardness are softness, little hardness, middle hardness and strong hardness; the values of Time length to prepare are no time, 1 hours, several hours, 1day and ≥1week. Hence, we firstly use discretization values 1,2,3… to represent them.

3.1. Factor Impact Analysis

Analysis results based on Pearson correlation are shown in Table 1. Apparently, age and time length to prepare are positive correlative with taking drugs which means older the male is, more potential the male goes to take medicine. Frequency of sexual behaviors and erection hardness are negative correlative with taking drugs which means softer the hardness is, more potential the male goes to take drugs.

Table 1. Pearson correlation analysis

		Age	Frequency	Hardness	Time to prepare
Whether to take drugs	Pearson Correlation	0.094	-0.102	-0.205	0.116
	Sig. (2-tailed)	<0.01	<0.01	<0.01	<0.01

Analysis results based on logistic regression are shown in Table 2. Obviously, hardness has biggest impact for determining taking drug potential as common sense, and

frequency has smallest impact among factors. Moreover, impact trend of each factor is compatible with correlation presented in above subsection. Compared with Pearson correlation analysis, logistic regression considering confounding effect make difference among factor impacts more evident.

Table 2. Logistic regression analysis

					95% C.I.for EXP(B)	
	B	**S.E.**	**Sig.**	**Exp(B)**	**Lower**	**Upper**
Age	0.168	0.010	<0.01	1.183	1.159	0.168
Frequency	-0.059	0.006	<0.01	0.943	0.932	-0.059
Hardness	-0.513	0.013	<0.01	0.599	0.583	-0.513
Time to prepare	0.217	0.009	<0.01	1.243	1.220	0.217

3.2. Decision Rule Learning

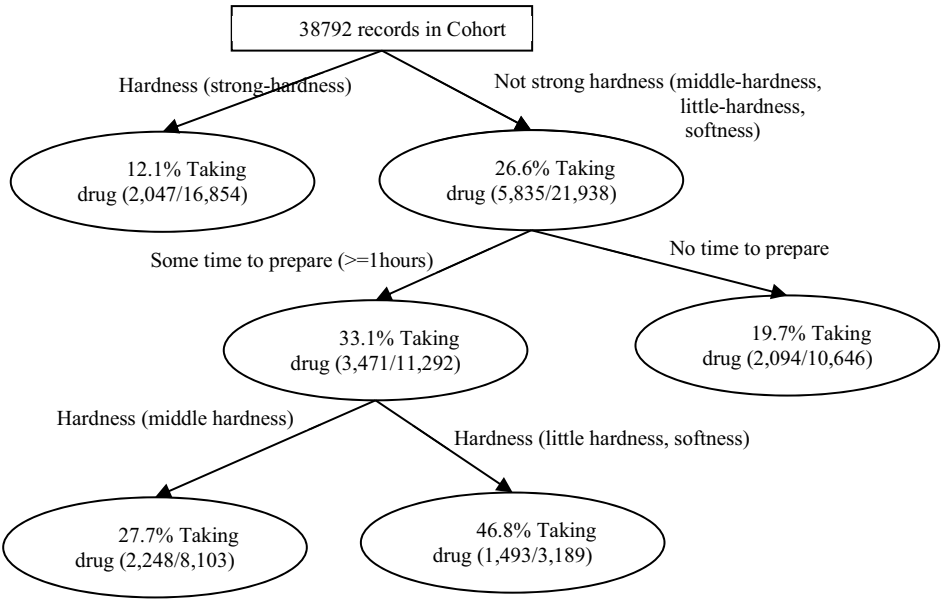


Figure 1. Predictors of taking drugs and potential stratification for the train cohort

Fig. 1 depicts the final tree generated by the CART analysis along with the questionnaire data for each child node of this tree. These branch points permit user stratification into four groups: high potential, intermediate potential-1, intermediate potential-2 and low potential.

Of the 4 variables (Age, Frequency, Hardness, and Time length to prepare) evaluated, the CART method firstly identifies Hardness factor as the best single discriminator between taking drugs and not taking drugs to split data into two nodes, left leaf node with strong-hardness as low potential and right child node with not strong-hardness. The next best predictor of taking drugs in not strong-hardness node is time length to prepare. The node with not strong-hardness is split into two new nodes, right leaf node with no time to prepare as high potential and left child node with some time to prepare. For the node with some time to prepare and not strong-hardness, a higher granularity level of hardness (middle harness, and little hardness, softness) is used to furthermore partition the data

into two nodes, left leaf node with middle hardness as intermediate potential-1 and right leaf node with little hardness or softness as intermediate potential-2. The interesting founds is that the patients with little hardness or softness has highest possibilities to take drugs rather than middle hardness.

Table 3. Taking drugs OR values between potential groups in test cohort

					95% C.I.for EXP(B)	
Low Potential (Reference)	B	S.E.	Sig.	Exp(B)	Lower	Upper
intermediate potential-2	0.572	0.067	<0.01	1.772	1.554	2.021
intermediate potential-1	0.980	0.067	<0.01	2.664	2.335	3.040
high potential	1.808	0.085	<0.01	6.098	5.159	7.209
					95% C.I.for EXP(B)	
intermediate potential-2 (Reference)	B	S.E.	Sig.	Exp(B)	Lower	Upper
intermediate potential-1	0.408	0.069	<0.01	1.503	1.313	1.721
high potential	1.236	0.087	<0.01	3.441	2.904	4.078
					95% C.I.for EXP(B)	
intermediate potential-1 (Reference)	B	S.E.	Sig.	Exp(B)	Lower	Upper
high potential	0.828	0.087	<0.01	2.289	1.931	2.713

Table 3 summarizes the OR results in these 4 potential groups based on test data. The OR value between the high-potential and low-potential groups was 6.098 (95% confidence interval, 5.159-7.209), with statistically significant differences in taking drug potential detected between all potential groups. Although additional nodes involving additional variables could be generated, they offered little incremental discrimination.

Conclusions

In this paper, we use statistics and data mining approaches to analyze correlation between male sexual behaviors and use of erectile dysfunction drugs. In the future, the approaches are general and can be adapted to large number of factors analysis tasks.

References

[1] I.P. Stolerman, L.H. Price. *Encyclopedia of Psychopharmacology*, Springer Berlin Heidelberg, 2010
[2] M. Braun, G. Wassmer, T. Klotz, B. Riefenrath. Mathers M, Engelmann U. Epidemiology of erectile dysfunction: results of the 'Cologne Male Survey'. *International Journal of Impotence Research* 12(2000), 305-311
[3] K. Zhang, W. Yu, Z.J. He, J. Jin. Help-seeking behavior for erectile dysfunction: a clinic-based survey in China. *Asian Journal of Andrology* 16(2014)
[4] K. Zhang, B. Xu, D.F. Liu, X.F. Wang, J.C. Zhu, J.J. Jiang H. Medical management of erectile dysfunction in aging males: Is it too late to treat? *Asian Journal of Andrology* 16(2014), 153-156
[5] M.H. Blanker, J.L. Bosch, F.P. Groeneveld et al, Erectile and ejaculatory dysfunction in a community-based sample of men 50–78 years old: prevalence, concern, and relation to sexual activity, *Urology* 57(2001), 763–8
[6] G.C. Fonarow. K.F. Adams, W.T. Abraham, et al, Risk stratification for in-hospital mortality in acutely decompensated heart failure: classification and regression tree analysis, *JAMA* 293(2005), 572-580.
[7] K. Pearson, Notes on regression and inheritance in the case of two parents, Proceedings of the Royal Society of London 58(1985), 240–242.

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-211

Automated Identification of National Health Survey Research Topics in the Academic Literature

Dean William YERGENS[a,b,1], Daniel James DUTTON[c] and Kirsten Marie FIEST[a,d,e]
[a] *Department of Critical Care Medicine, Cumming School of Medicine, University of Calgary, Canada*
[b] *Synthesis Research Inc., Calgary, Canada*
[c] *School of Public Policy, University of Calgary, Canada*
[d] *O'Brien Institute for Public Health, University of Calgary, Canada*
[e] *Hotchkiss Brain Institute, University of Calgary, Canada*

Abstract. National health surveys are routinely conducted to provide value data about a country's health status and the health services being consumed by the population. This information is used for surveillance, research, and the planning of healthcare services at local and national levels. Although these national health surveys are viewed as important resources for public and population health, there is limited information as to the type of research being conducted with these surveys. This study investigates, through the use of automated text data mining, an approach to identify and collate the type of academic literature being published using national health surveys.

Keywords. Text data mining, algorithm, epidemiology, literature review, national health survey

1. Introduction

National health surveys are routinely conducted to provide value data about a country's health status and the health services being consumed by the population. This information is used for surveillance, research, and the planning of healthcare services at local and national levels. Because of the availability and comprehensiveness of these surveys, this data is routinely analyzed by various academic and government institutions with the results being disseminated through the academic literature. Many countries engage in conducting national health surveys, as illustrated in Table 1, which not only allows for national comparisons, but for international comparisons as well.

The type of epidemiology and health services research being conducted with these national health survey datasets has not been described in the literature to our knowledge. Previous studies have examined the use of statistical methods in the biomedical literature[1], as well as individual journals[2]. Text data mining has also been

[1] Corresponding Author: Dean William Yergens, Department of Critical Care Medicine, Faculty of Medicine, University of Calgary, Calgary, Alberta, Canada; E-mail: dyergens@ucalgary.ca.

applied to examine the statistical methods utilized in the published medical literature[3] and within a Canadian national health survey[4].

The objective of this study was to develop an automated approach for determining the type of epidemiology and health services research being conducted using national health surveys and published in the academic literature, based strictly upon identifying specific keywords and phrases found in the paper's title.

Table 1. Examples of National Health Surveys

Survey	Country	Description
CCHS	Canada	Canadian Community Health Survey (CCHS) is a cross sectional survey, conducted since 2001, that collects health status, health care services utilization and health determinants data of the Canadian population. [5]
BRFSS	United States	Behavioral Risk Factor Surveillance System (BRFSS) is a health survey, conducted since 1984, collects health related risk behaviors, chronic health conditions, and use of healthcare services for the United States (US) population.[6]
NHANES	United States	National Health and Nutrition Examination Survey (NHANES), conducted since the early 1960s, assesses the health and nutritional status of adults and children in the US.[7]
KNHANES	South Korea	Korean National Health and Nutrition Examination Survey (KNHANES) is conducted to evaluate the health and nutritional status of the South Korean population. [8]

2. Methods

2.1. Literature Search Strategy

Literature searches were conducted for each of four national health surveys appearing in Table 1 using the PubMed bibliographical database. For the CCHS, our search strategy consisted of the phrase "Canadian Community Health Survey"; BRFSS consisted of "Behavioral Risk Factor Surveillance System"; NHANES consisted of the term "NHANES" or the phrase "National Health and Nutrition Examination Survey" with both of them excluding the terms "Korea*"; and the KNHANES search consisted of term "KNHANES" or the phrase "Korean National Health and Nutrition Examination Survey". All of the searches were limited to title and abstract only and were conducted on October 24, 2016. We did not assess if the national health survey was the major topic of the paper or was only referenced.

2.2. Data Management and Custom Software

All of the references were imported into a custom written Java-based literature reference management program (Synthesis). This software was created by DWY and is described in more detail elsewhere[4]. Synthesis is built upon the open-source Apache Lucene database and has the ability to manage textual documents for collating, managing, and performing Boolean queries based upon the imported references in the Lucene database. The Synthesis software is capable of taking a text definition file based upon keywords or phrases, Boolean operators, wildcards, and proximity searching and tag every reference that meets the user-defined criteria.

2.3. Topic Algorithm Development

To determine the research topics, DWY and KMF used the CCHS references to identify the main categories. Main categories consisted of keywords, phrases, and basic algorithms. This was an iterative process, which involved looking for commonly used words in the titles of the CCHS references and determining whether it was a suitable candidate to include within the main categories. This is described below.

To aid in the identification of commonly occurring words, a dynamic Word Cloud included in Synthesis was utilized. Once a topic/concept of interest was identified, a combination of Boolean logic (AND, OR, NOT) and wildcards was used to construct a statement that could automatically identify and tag the concept within Synthesis. An example of a statement defining the concept of Determinants would be: 'title:determinants AND NOT title:"social determinants"'.

Once a concept was identified and determined suitable for inclusion, it was grouped into a higher-level main category. It should be noted that many frequently appearing words in the title were not suitable to be included as they were either too general and often did not reflect a topic categorization. Titles where the algorithm produced no categories were then marked as 'Unclassified'. We did encounter several references in the Unclassified category that could benefit from more in depth analysis and rules. An example of this were references that simply had the Outcome and the Exposure as the title (e.g. Smoking and oral health status) which would indicate belonging to the Association category. However, to categorize these kinds of references, a list of potential domain area variables (e.g. smoking, oral health, etc.) would need to be constructed which we determined was outside of the scope of this paper.

Table 2. Category Definitions

Main Category	Category Concepts
Characteristics	Characteristics, Epidemiology, Determinants, Factor, Comorbidity, Consumption, Behavio(u)r, Burden, Unmet, Inequality, Inequity, Profile, Classification, Descriptive, Among
Association	Association, Between, Relationship, Differences, Comparison, Variation, Correlation, Disparities, Link
Estimates	Estimates, Prevalence, Incidence, Occurrence, Adjusted
Surveillance	Surveillance, Trends, Increase, Decrease, Change, Pattern, Update, Incremental, Screening, Rate
Risk	Risk
Utilization	Utilization, Usage, Access, Services, Treatment
Prediction	Prediction, Forecast, Impact, Adherence, Determinant
Evaluation	Evaluation, Validation, Accuracy, Reliability
Implementation	Implementation, Application, Planning, Management, Recommendation
Methodology	Methodology, Algorithm, Derive, Design, Develop
Spatial	Spatial, Geographical, Map
Unclassified	No categories identified

In total, eleven main categories were identified: Association, Characteristics, Estimates, Surveillance, Risk, Utilization, Implementation, Validation, Prediction, Methodology, and Spatial. Each of these main categories consisted of a variety of associated keywords/concepts as a main category could represent several associated concepts. For example, the Utilization main category consisted of derivatives of the following keywords: utilization, usage, access, services, and treatment. A list of the main categories and their associated sub-categories can be found in Table 2. It should be noted that a paper may be tagged with more than one category.

3. Results

Four separate literature searches were conducted. The search for the CCHS dataset resulted in 996 references, BRFSS 2289 references, NHANES 8286 references, and KNHANES 986 references. The 11 main concept algorithm definition file was applied within Synthesis to each of the four datasets (see results in Table 3).

The two most frequent main categories across all datasets were Characteristics and Association. Characteristics was identified in 25.0%, 34.1%, 19.0%, and 17.3% of the CCHS, BRFSS, NHANES, and KNHANES datasets, while Association was identified in 20.1%, 15.3%, 22.7%, and 33.4% of the datasets. The Estimates, Surveillance, and Risk main categories were the next most frequent groups, accounting for 7.5%, 7.5%, and 5.3% of the CCHS references, 10.2%, 11.7%, and 6.1% of the BRFSS references, 8.3%, 7.8%, and 8.9% of the NHANES references, and 8.9%, 9.0%, and 9.9% of the KNHANES references. The next natural grouping of main categories based upon percentages identified in the four datasets consisted of Utilization, Prediction, Evaluation, Implementation, Methodology, and Spatial. These main categories were represented in the low single digit percentage of all categories amongst the four datasets. References where a category could not be identified were labeled as Unclassified and percentages across the databases ranged from 9.7% (BRFSS), 12.9% (KNHANES), 18.5% (CCHS), to 22.6% (NHANES).

Table 3. Results from Topic Algorithm

Main Category	CCHS	BRFSS	NHANES	KNHANES
Characteristics	345 (25.0%)	1214 (34.1%)	2107 (19.0%)	247 (17.3%)
Association	278 (20.1%)	545 (15.3%)	2516 (22.7%)	478 (33.4%)
Estimates	104 (7.5%)	362 (10.2%)	914 (8.3%)	127 (8.9%)
Surveillance	103 (7.5%)	417 (11.7%)	867 (7.8%)	128 (9.0%)
Risk	73 (5.3%)	217 (6.1%)	989 (8.9%)	141 (9.9%)
Utilization	77 (5.6%)	166 (4.7%)	199 (1.8%)	23 (1.6%)
Prediction	54 (3.9%)	102 (2.9%)	419 (3.8%)	47 (3.3%)
Evaluation	27 (2.0%)	55 (1.5%)	204 (1.8%)	16 (1.1%)
Implementation	15 (1.1%)	57 (1.6%)	138 (1.2%)	17 (1.2%)
Methodology	27 (2.0%)	32 (0.9%)	200 (1.8%)	19 (1.3%)
Spatial	22 (1.6%)	43 (1.2%)	15 (0.1%)	2 (0.1%)
Unclassified	255 (18.5%)	346 (9.7%)	2499 (22.6%)	185 (12.9%)

4. Conclusion

This study provides an approach to using text data mining for categorizing research topics of national health surveys based upon the titles of academic publications. This study reports on four commonly used national health surveys from multiple countries and finds that title topic categorizations are relevantly consistent across all of datasets. This indicates that topic definitions could be applied to other health surveys outside of the CCHS, for which it was originally developed.

This study identifies three natural boundaries in the research being produced from national health surveys. The first grouping includes Characteristics and Associations, which account for roughly 40-50% of the research being published. The second group consists of Estimates, Surveillance, and Risk, accounting for 20-30% of the publications. The third group, includes the Utilization, Prediction, Evaluation,

Implementation, Methodology, and Spatial main categories, which account for a small number of publications from these surveys.

The high percentage of references in the Characteristic and Association main category is expected. As national health surveys are readily available datasets, they provide a cost effective and timely solution for conducting much needed research in many different areas. The description of patient populations and associations between differing variables of interest is an important aspect of epidemiology and its application to population and public health. The finding that there were a low number of publications in the Implementation category is interesting, and we wonder if this is the result of actionable initiatives not being commonly reported in the academic literature. Knowing this information provides the opportunity to help guide future health policy into which areas should be strengthen and identifying gaps in the research.

There are several limitations in this study. First, the main categories identified are most likely only applicable to national health surveys or domains with a public or population health focus. We anticipate that other research areas have their own unique set of terminology and focus. An example of this could be clinical medicine, where survival analysis and outcome research may be more prevalent and require new categories. Second, we only analyzed titles to determine the research topic being investigated. Future research should also examine the abstract and full-text of each publication which could provide additional information to aid in categorizing the research topics.

References

[1] M. Scotch, M. Duggal, C. Brandt, Z. Lin Z, R. Shiffman, Use of statistical analysis in the biomedical informatics literature, *J Am Med Inform Assoc* Jan-Feb 17(1) (2010), 3-5.

[2] P.J. Becker, E. Viljoen, L. Wolmarans, C.B. IJsselmuiden, An assessment of the statistical procedures used in original papers published in the SAMJ during 1992, *S Afr Med J* Sep 85(9) (1995), 881-4.

[3] C. Meaney, R. Moineddin, T, Voruganti, M.A. O'Brien, P. Krueger, F. Sullivan, Text mining describes the use of statistical and epidemiological methods in published medical research, J Clin Epidemiol Jun 74 (2016), 124-32.

[4] D.W. Yergens, D.J. Dutton, S.B. Patten, An overview of the statistical methods reported by studies using the Canadian community health survey, *BMC Med Res Methodol* Jan 25 (2014), 14:15.

[5] Statistics Canada, Canadian Community Health Survey (CCHS), webpage: http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&Id=3359, Accessed Nov 6, 2016.

[6] Centers for Disease Control and Prevention, Behavioral Risk Factor Surveillance System (BRFSS), webpage: http://www.cdc.gov/brfss/, Accessed Nov 6, 2016.

[7] Centers for Disease Control and Prevention, National Health and Nutrition Examination Survey, webpage: http://www.cdc.gov/nchs/nhanes/, Accessed Nov 6, 2016.

[8] Korean National Health & Nutrition Examination Survey, Survey Overview, Webpage: https://knhanes.cdc.go.kr/knhanes/eng/sub01/sub01_02.do, Accessed Nov 6, 2016.

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-216

Prevalence Estimation of Protected Health Information in Swedish Clinical Text

Aron HENRIKSSON[a,1], Maria KVIST[a,b] and Hercules DALIANIS[a]
[a] *Department of Computer and Systems Sciences, Stockholm University, Sweden*
[b] *Department of Laboratory Medicine, Karolinska Institutet, Sweden*

Abstract. Obscuring protected health information (PHI) in the clinical text of health records facilitates the secondary use of healthcare data in a privacy-preserving manner. Although automatic de-identification of clinical text using machine learning holds much promise, little is known about the relative prevalence of PHI in different types of clinical text and whether there is a need for domain adaptation when learning predictive models from one particular domain and applying it to another. In this study, we address these questions by training a predictive model and using it to estimate the prevalence of PHI in clinical text written (1) in different clinical specialties, (2) in different types of notes (i.e., under different headings), and (3) by persons in different professional roles. It is demonstrated that the overall PHI density is 1.57%; however, substantial differences exist across domains.

Keywords. electronic health records, protected health information, de-identification, natural language processing, predictive modeling

1. Introduction

Healthcare produces an abundance of data that is stored in electronic health record (EHR) systems. The secondary use of EHR data, which describes the health conditions and treatments of patients over time, holds much promise in facilitating medical research and epidemiological activities; EHR data can also be exploited for providing clinical decision support at the point of care. However, this requires that privacy-preserving measures, such as de-identification, are taken. Automatic de-identification of EHR data includes the detection and obscuring of sensitive information in clinical notes. The US Health Insurance Portability and Accountability Act (HIPAA) defines 18 types of protected health information (PHI) that should be obscured for EHR data to be considered de-identified [1]. In recent years, there has been a surge in research efforts to construct automatic de-identification tools [2,3], many of which rely on machine learning and manually annotated corpora for identifying PHI in clinical notes.

In this study, we seek to estimate the prevalence of PHI in Swedish clinical text. In particular, we want to uncover if differences exist in the distribution of PHI – both generally and with respect to specific PHI classes – across different types of notes. That this may, in fact, be the case is substantiated by the knowledge that one writes differently in different clinical specialties and professional roles [4]. The findings from this study are also intended to inform future development of automatic de-identification systems.

[1] Corresponding author: aron.henriksson@dsv.su.se

Previous research has attempted to estimate PHI prevalence based on small samples of annotated data and covered only a few types of clinical notes. An early study based on a sample of nursing notes in MIMIC-II – an EHR database which has indeed been de-identified and made publicly available for research – revealed that around 0.5% of all tokens were instances of PHI [5]. In another study, PHI density in a diverse set of clinical domains was found to be 2.9% and name density 1% [6]. In discharge summaries, the PHI density amounted to around 3.6% [7]. A similar PHI density was found in the 2014 i2b2/UTHealth corpus, comprising health records of diabetic patients [8]. In one study, the distribution of PHI classes across different types of notes, i.e. written under different headings, was described: the most prevalent PHI types were dates, names and phone numbers, while the note types with the highest PHI density were *Discharge Summary*, *Outpatient Consult*, and *Admission History and Physical* [9]. A few similar studies have been performed on non-English languages. In a French corpus comprising clinical notes of various types from a range of specialties, PHI density was as high as 11% [10], while a study of Danish clinical text revealed a PHI density of around 1.8% [11].

2. Methods & Materials

This study seeks to estimate the prevalence of PHI in Swedish clinical text and to learn if differences therein exist between types of clinical text. As manual annotation is cumbersome, we estimate PHI prevalence by applying a predictive model that has been trained on an existing PHI corpus to a larger unannotated corpus of clinical text. Both the annotated corpus [12] and the data extracted for this study are from the Stockholm EPR Corpus[2] [13], which contains health records from Karolinska University Hospital.

The annotated PHI corpus comprises 100 health records from five different clinics (Neurology, Orthopedics, Infection, Dental Surgery, and Nutrition) in 2008. The corpus contains a total of 198,466 tokens (0.1 types/token) and 4,220 annotated PHI instances. The PHI density is 2.13% and the class distribution is as follows: Health Care Unit (23.9%), First Name (21.7%), Last Name (21.5%), Date Part (16.6%), Full Date (8.7%), Location (3.3%), Phone Number (3.2%), Age (1.2%). The average sentence length is 8.9 (± 6.4) tokens. Approximately 13.7% of all sentences include a least one PHI mention, while, on average, there are 0.19 (± 0.55) PHI mentions per sentence.

This manually annotated PHI corpus was used for training a predictive model. To that end, a linear-chain CRF [14] was used that, in addition to being dependent on the input features, is also dependent on the previous and subsequent output variable. The following features were used: (a) token, (b) lemma, (c) part of speech, (d) capitalization, (e) digit, (f) compounds, (g) dictionary matching against SNOMED CT, MeSH etc. The same features were used in previous studies on named entity recognition in Swedish clinical text [15,16] (see [15] for more details). As is common for sequence labeling tasks, IOB-encoding of class labels was used, which indicates whether a token is at the beginning (B), inside (I) or outside (O) a given named entity mention. 10-fold cross-validation was carried out when tuning the CRF hyperparameters: two forms of regularization (L1/L2), the c-value governing the balancing between underfitting and overfitting, and the window size, which determines to what extent dependencies should be modeled between input features and output variables. Considered c-values were 2^x,

[2] This research has been approved by the Regional Ethical Review Board in Stockholm (2012/834-31/5).

where x ∈ {-2,-1,0,1,2,3,4,5}, while the following symmetric window sizes were explored: 1+1, 2+2, 3+3, 4+4.

A predictive model was then trained on the entire annotated PHI corpus using the best-observed set of hyperparameters and subsequently applied on various subsets of the unannotated corpus, comprising all clinical notes from a single year: 2009. PHI prevalence was estimated and compared along three dimensions: (1) specialty, i.e., notes from units belonging to different clinical practices (geriatrics, oncology, orthopedics), (2) note types, i.e., notes written under different headings (admission, day, discharge), and (3) professions, i.e., notes written by persons in different professional roles (physicians, nurses, physiotherapists). Since these subcorpora are not equal in size, we use a normalized metric, PHI density, to quantify prevalence: this is defined as the number of PHI mentions divided by the total number of tokens in the corpus.

3. Results

The results of the parameter optimization are shown in Figure 1. The trend was the same irrespective of whether F_1-scores were micro- or macro-averaged over classes: L2 regularization led almost invariably to higher performance, with higher c values favored in comparison to L1 regularization. The best results (precision: 92.65%, recall: 81.29%, F_1: 0.87) were obtained using a narrow context window (1+1) and a c value of 16.

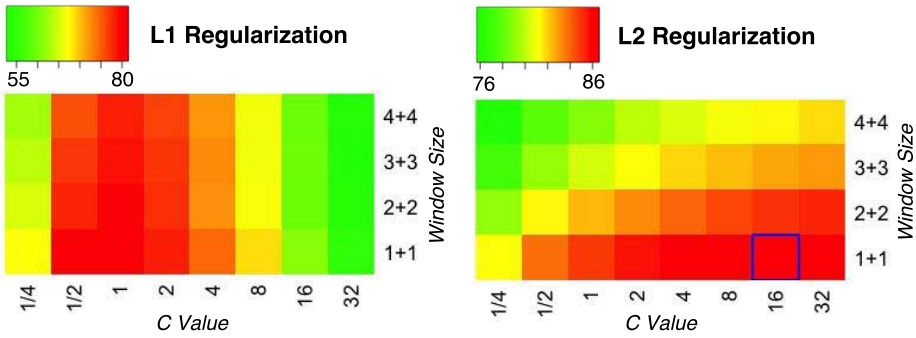


Figure 1. Parameter optimization results with 10-fold cross-validation on the training set

A predictive model was then trained and applied to various subsets of clinical notes from a different year. Descriptive statistics of the sub-corpora along with PHI density estimates are shown in Table 1. The average sentence length is 9.1 tokens, with no domain standing out from the others. There are, however, differences in type-token ratios, with larger lexical variation observed in notes written by physicians (0.007) and nurses (0.008) compared to physiotherapists (0.015); there is less lexical variation in geriatrics (0.016) than in oncology (0.011) and orthopedics (0.010). Interesting to note is that the overall type-token ratio is as low as 0.004. On average across domains, the PHI density is 1.57%. There is an average of 0.14 PHI tokens per sentence and around 10% of sentences contain at least one PHI instance. With respect to different specialties, there is a fairly substantial difference in PHI density between orthopedics (1.50%), on the one hand, and geriatrics (2.12%) and oncology (2.05%), on the other. Looking at specific PHI classes, there are relatively fewer dates in the notes produced in orthopedics;

geriatrics mentions relatively more first names, ages and phone numbers; oncology generally stands out less, but there does seem to be a propensity for writing dates. When it comes to different types of clinical notes, the observed PHI distribution is even more skewed: admission notes comprise the least amount of PHI (1.00%), discharge notes the most (2.94%), and day notes somewhere in the middle (1.64%). Names and health care units are particularly prevalent in discharge notes in comparison to the other note types. Regarding notes written by persons in different professions, differences in PHI density are somewhat smaller, with physiotherapist notes exhibiting the least amount of PHI (0.97%), followed by physician notes (1.43%) and nurse notes (1.66%). Nurses seem to mention names, dates and locations to a greater extent than do the other professions.

Table 1. PHI density estimates across various types of clinical text

	Specialty			**Note**			**Profession**		
	Geriatrics	Oncology	Orthopedics	Admission	Day	Discharge	Dr	Nurse	Physio
Sentence Length	9.9 ± 8.5	9.5 ± 7.0	8.7 ± 6.5	8.4 ± 7.1	9.5 ± 6.9	9.6 ± 8.8	9.5 ± 7.6	8.1 ± 5.8	10.4 ± 8.6
Type:Token	0.016	0.010	0.011	0.015	0.014	0.017	0.007	0.008	0.015
PHI Density (%)	2.12	2.05	1.50	1.00	1.61	2.94	1.43	1.66	0.97
First Name	0.55	0.32	0.33	0.11	0.29	0.58	0.29	0.37	0.31
Last Name	0.44	0.39	0.45	0.17	0.35	0.68	0.36	0.44	0.22
Age	0.04	0.01	0.02	0.05	0.01	0.06	0.03	0.00	0.01
Health Care Unit	0.32	0.30	0.24	0.25	0.20	0.47	0.26	0.16	0.11
Location	0.07	0.07	0.06	0.10	0.06	0.08	0.09	0.58	0.04
Full Date	0.30	0.27	0.17	0.10	0.12	0.59	0.17	0.14	0.09
Date Part	0.32	0.68	0.20	0.22	0.56	0.43	0.22	0.44	0.16
Phone Number	0.08	0.02	0.03	0.00	0.03	0.05	0.01	0.04	0.04

4. Discussion

This study sought to estimate the prevalence of PHI in Swedish clinical text and investigate differences across various types of notes. The amount of sensitive information (1.57%) is comparable to previous reports on other languages, although numbers range from 0.5% to 11%. It is, however, problematic to compare these directly as different PHI classes and definitions are used. In contrast to previous work, we looked specifically at different types of notes written in different specialties and professions. This revealed some notable differences in PHI density, primarily when comparing admission, day and discharge notes. The highest PHI density was observed in discharge notes: almost 20% of the sentences contained at least one PHI instance. Many plausible explanations can be found for the observed differences in PHI density: names, especially surnames are, e.g., prevalent in discharge summaries in part because physicians involved in the healthcare process are typically mentioned; that physicians mention healthcare units and nurses mention locations, respectively, can be attributed to the fact that physicians tend to write about the healthcare process both within the hospital and with general practitioners, while nurses need also to coordinate with the outside world.

Knowing about differences in PHI density is useful for development of de-identification systems, e.g. when creating training data for machine learning. In future work, we plan to evaluate the predictive performance on different types of notes in order to assess to what extent domain adaptation may be necessary. In comparison to previous

studies, in which small annotated corpora were used, we proposed an alternative way of estimating PHI prevalence by using predictive modeling. While highly efficient, there are limitations in terms of reliability as the estimates are dependent on the performance of the predictive model. Here, we obtained an F_1-score of 0.87, cross-validated on the training set, outperforming previous models trained on the same corpus [12,17,18]; however, in future work, we also need to determine the performance on the target domains.

References

[1] HIPAA 2003. Health Insurance Portability and Accountability (HIPAA), Privacy Rule and Public Health Guidance, 2003, From CDC and the U.S. Department of Health and Human Services, 2016. Accessed 2016-10-18.

[2] Özlem Uzuner, Yuan Luo, and Peter Szolovits. Evaluating the state-of-the-art in automatic de-identification. *Journal of the American Medical Informatics Association*, 14(5):550–563, 2007.

[3] Stephane Meystre, Jeffrey Friedlin, Brett South, Shuying Shen, and Matthew Samore. Automatic de-identification of textual documents in the electronic health record: a review of recent research. *BMC medical research methodology*, 10(1):70, 2010.

[4] Kelly Smith, Beata Megyesi, Sumithra Velupillai, and Maria Kvist. Professional language in Swedish clinical text: Linguistic characterization and comparative studies. *Nordic Journal of Linguistics*, 37(02):297–323, 2014.

[5] Margaret Douglass, Gari D Clifford, Andrew Reisner, George B Moody, and Roger G Mark. Computer-assisted de-identification of free text in the mimic ii database. *Computers in Cardiology*, pages 341–344, 2004.

[6] David A Dorr, WF Phillips, Shobha Phansalkar, Shannon A Sims, and John Franklin Hurdle. Assessing the difficulty and time cost of de-identification in clinical narratives. *Methods of information in medicine*, 45(3):246–252, 2006.

[7] Özlem Uzuner, Tawanda C Sibanda, Yuan Luo, and Peter Szolovits. A de-identifier for medical discharge summaries. *Artificial intelligence in medicine*, 42(1):13–35, 2008.

[8] Amber Stubbs and Özlem Uzuner. Annotating longitudinal clinical narratives for de-identification: The 2014 i2b2/uthealth corpus. *Journal of biomedical informatics*, 58:S20–S29, 2015.

[9] David Hanauer, John Aberdeen, Samuel Bayer, Benjamin Wellner, Cheryl Clark, Kai Zheng, and Lynette Hirschman. Bootstrapping a de-identification system for narrative patient records: cost-performance tradeoffs. *International journal of medical informatics*, 82(9):821–831, 2013.

[10] Cyril Grouin and Aurélie Névéol. De-identification of clinical notes in French: towards a protocol for reference corpus development. *Journal of biomedical informatics*, 50:151–161, 2014.

[11] Kostas Pantazos, Soren Lauesen, and Soren Lippert. De-identifying an EHR database - anonymity, correctness and readability of the medical record. In *Medical Informatics Europe*, pages 862-866, 2011.

[12] Hercules Dalianis and Sumithra Velupillai. De-identifying Swedish clinical text-refinement of a gold standard and experiments with Conditional random fields. *J. Biomedical Semantics*, 1:6, 2010.

[13] Hercules Dalianis, Martin Hassel, Aron Henriksson, and Maria Skeppstedt. Stockholm EPR Corpus: a clinical database used to improve health care. In *Swedish Language Technology Conference*, 2012.

[14] John Lafferty, Andrew McCallum, and Fernando CN Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. 2001.

[15] Maria Skeppstedt, Maria Kvist, Gunnar H Nilsson, and Hercules Dalianis. Automatic recognition of disorders, findings, pharmaceuticals and body structures from clinical text: an annotation and machine learning study. *Journal of biomedical informatics*, 49:148–158, 2014.

[16] Aron Henriksson, Maria Kvist, Hercules Dalianis, and Martin Duneld. Identifying adverse drug event information in clinical notes with distributional semantic representations of context. *Journal of biomedical informatics*, 57:333–349, 2015.

[17] Aron Henriksson, Hercules Dalianis, and Stewart Kowalski. Generating features for named entity recognition by learning prototypes in semantic space: The case of de-identifying health records. In *International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 450–457, 2014.

[18] Aron Henriksson. Learning multiple distributed prototypes of semantic categories for named entity recognition. *International journal of data mining and bioinformatics*, 13(4):395–411, 2015.

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-221

The Effects of Heterogeneity in the Comparative Effectiveness of Individual Treatments in Randomised Trials

Paraskevi PERICLEOUS[a,1], Tjeerd van STAA[a] , Matthew SPERRIN[a] and on behalf of GetReal Work Package 2
[a] *Farr Institute, Faculty of Biology, Medicine and Health, University of Manchester, Manchester Academic Health Science Centre*

Abstract. In some randomised trials, the new treatment can be compared with usual care which can include multiple treatments and result in a heterogeneous control group. In this paper, we use simulation to assess the performance of various statistical methods to infer the individual effects of the various control treatments. These methods include inverse Probability Weighting, Doubly Robust Inverse Probability Weighting, Propensity Score, Disease Risk Score, Standardization and Multivariable Logistic Regression. Different scenarios were tested including unmeasured heterogeneity with or without confounding. The methods perform well when heterogeneity and confounding are both fully captured; however, for the scenarios where heterogeneity is not fully captured this leads to biased effect estimates, particularly where there is also unobserved confounding. Thus, leading to potentially misleading comparative effectiveness of individual treatments.

Keywords. Heterogeneity, confounding, unobserved, trials

1. Introduction

Randomised controlled trials (RCTs) are often performed under artificial conditions in narrowly selected patient group, while their results are often generalised to more diverse patient groups in routine practice [1] [2] [3]. Additionally, RCTs are expensive, time-consuming, and often fail to meet their recruitment targets [4] [5] [6]. Pragmatic trials, on the other hand, are performed in the real clinical practice environment [1] [4] and typically do not have strict entry criteria [1] [7]. They can involve a comparison of a new treatment with usual care (rather than placebo) [5]. There are often multiple treatments that are used in usual care. While the comparison between the new and the usual care group will be unbiased due to the randomisation, the specific comparisons with the individual control treatments may be more difficult due to heterogeneity in the patient groups using these treatments, which may lead to confounding. Typical RCTs do not include heterogeneity in the control arm, while the pragmatic approach in this paper does.

[1] Corresponding author, Vaughan House, University of Manchester, M13 9GB;
E-mail: paraskevi.pericleous@manchester.ac.uk.

The aim of this paper is to compare different methods for adjusting for measured confounding in the presence of unmeasured heterogeneity that may or may not include confounding. We have chosen to do this via simulations and the methods that will be compared are multivariable logistic regression, inverse probability weighting, propensity score matching, disease risk score adjustment, direct standardisation and doubly robust inverse probability weighting; these are some of the commonly used methods to correct for measured confounding [8] [9]. These methods are used to predict the potential outcome if all patients were treated with the same treatment. The effect that treatment has to this potential outcome is called causal effect. This effect could either come from a marginal or a conditional model resulting in 'marginal causal effect' and 'conditional causal effect' respectively. To find the causal effect there has to be exchangeability and collapsibility. Exchangeability means that the potential outcome for each of the treatments and the treatment itself are independent and this is ensured by the randomisation [8]. When adjusting for a confounder C, if there is no change in the measure then there is collapsibility over C, if there is a change then the measure is non-collapsible over C [10]. Scenarios of fully observed heterogeneity and partially measured confounding have a non-exchangeability problem as the observed confounder is no longer sufficient for confounding adjustment [8] and a non-collapsibility problem as the outcome depends on the unobserved confounder [10]. Non-collapsibility is also a problem for the unmeasured heterogeneity, but since the confounding in this case is sufficient for the adjustment, this is no problem for the marginal causal effect [10].

2. Methods

Let Z denote the treatment allocation with $Z = 0$ for the new treatment and $Z = 1{,}2$ for two control treatments, where $Z = 1$ is assumed to be the baseline control treatment. Let n be the number of patients taking part in the trial. $n = \sum_{k=0}^{2} n_k$, where n_k denotes the number of patients taking treatment k. Let Y be a binary outcome (e.g. dead or alive), and C and U be additional variables (observed and unobserved respectively).

C is simulated as a binary variable, Bernoulli(0.5), while U is simulated from a normal distribution $N(0, \sigma^2)$. Patients are randomised to receive the new treatment with probability 0.5, and we assume no refusals. Otherwise, patients receive their usual care, according to the model:

$$\text{logit}\,(P[Z = 2|C, U]) = \alpha_0 + \alpha_1 C + \alpha_2 U + \alpha_3 CU, \quad (Z = 1 \text{ otherwise}). \qquad (1)$$

The outcome Y is simulated according to

$$\text{logit}\,(Y|Z, C, U) = \beta_0 + \beta_1 I[Z = 0] + \beta_2 I[Z = 2] + \beta_3 C + \beta_4 U, \qquad (2)$$

where $I[.]$ denotes the indicator function.

Each scenario of interest is described in Table 1. 10000 patients and 10000 replications will be used for all scenarios. Our key interest is in the bias in the estimation of β_1 and β_2, the log-odds of the new treatment with respect to the baseline control treatment and log-odds of the second control treatment with respect to the baseline control treatment, respectively.

Table 1. Simulation Scenarios

Scenario	Parameter settings
(a) No unmeasured heterogeneity or confounding	$\alpha_0=\alpha_2=\alpha_3=\beta_0=\beta_4=0$, $\alpha_1=\beta_1=\beta_2=\beta_3=1$
(b) Unmeasured heterogeneity with confounding	$\alpha_0=\beta_0=0$, $\alpha_1=\alpha_2=\alpha_3=\beta_1=\beta_3=1$, $\sigma=1$. $\beta_2=\beta_4=1$ or $\beta_2=\beta_4=-1$
(c) Unmeasured Heterogeneity without confounding	$\alpha_0=\alpha_2=\alpha_3=\beta_0=0$, $\alpha_1=\beta_1=\beta_2=\beta_3=\beta_4=1$, $\sigma=0.7$

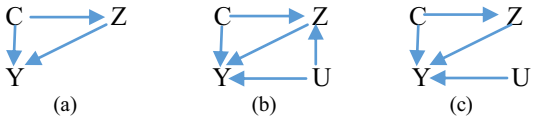


Figure 1. Causal diagrams for the three scenarios. C, Z and Y, denote an observed confounder, treatment and binary outcome respectively. U denotes an unmeasured variable. Panel (a) represents 'no unmeasured heterogeneity or confounding' scenario; panel (b) the 'unmeasured heterogeneity with confounding' scenario; and panel (c) the scenario 'unmeasured heterogeneity without confounding'.

The methods used for adjusting for confounding are: (1) Multivariable logistic regression corrected for the confounder C; (2) propensity score method, which uses the probability of receiving a specific treatment given the observed confounder C as a continuous covariate with the other variables [9], (3) Disease Risk Score Adjustment method that uses the predicted probability of the event (alive or dead) given the observed covariates and used as a continuous covariate with the other variables [9], (4) Inverse Probability Weighting method that uses stabilised weights obtained from the propensity score [8] and used as weights without the confounders. (5) Doubly Robust Inverse Probability Weighting, like Inverse Probability Weighting but with the confounders in the outcome model [8]. (6) Standardisation, which is standardising the mean outcome to the confounder distribution [8].

3. Results

Let us denote the mean of all the estimates of β_1 and β_2 from each method as $\overline{\widehat{\beta_1}}$ and $\overline{\widehat{\beta_2}}$ respectively. Figure 2 represents the estimates of β_1 and β_2 for scenario (a). The first four methods estimate the conditional causal effect (Logistic, DRIPW, PS, DRS) and the final two estimate the marginal causal effect (IPW and Standardisation). All methods are estimating the conditional and marginal causal effects accurately. This was expected as the confounding and heterogeneity are fully captured. Figures 3 and 4 shows the estimates of β_1 and β_2 for scenario (b). In Figure 3, we have $\beta_2=\beta_4=1$. Here, all methods estimating the conditional causal effect find $\overline{\widehat{\beta_1}}=0.452$, except from DRIPW that finds $\overline{\widehat{\beta_1}}=0.458$. IPW and standardisation find $\overline{\widehat{\beta_1}}=0.444$ and 0.438, respectively. All methods estimating the conditional causal effect find the mean of $\overline{\widehat{\beta_2}}=0.016$, except for DRIPW, which gives a value of 0.003 (same as IPW). Standardisation estimates $\overline{\widehat{\beta_2}}=0.016$. All methods therefore underestimate the marginal and the conditional causal effect. For β_2, however, even though the actual value of both the marginal and conditional causal effects are positive, all the estimated values are close to zero with some even negative. Figure 4 again considers scenario (b), but with $\beta_2=\beta_4=-1$. Here, all methods estimating the conditional causal effect find $\overline{\widehat{\beta_1}}=1.203$, except from DRIPW that finds $\overline{\widehat{\beta_1}}=1.198$. IPW and Standardisation find $\overline{\widehat{\beta_1}}=1.141$ and

1.147, respectively. All methods estimating the conditional causal effect find $\overline{\overline{\beta_2}}$= 0.009, except from DRIPW which gives a value of 0.025. IPW and Standardisation find $\overline{\overline{\beta_2}}$= 0.024 and 0.008, respectively. All methods overestimate the marginal and the conditional causal effect. For β_2, however, even though the actual value of both marginal and conditional causal effect is negative, all the estimated values are close to zero with some even positive. Figure 5 represents the estimates of β_1 and β_2 for scenario (c). All methods estimating the conditional causal effect find $\overline{\overline{\beta_1}}$= 0.958, except from DRIPW that finds it $\overline{\overline{\beta_1}}$ = 0.955. IPW and Standardisation find $\overline{\overline{\beta_1}}$= 0.913 and 0.915, respectively. All methods estimating the conditional causal effect find $\overline{\overline{\beta_2}}$ = 0.957, except for DRIPW which gives a value of 0.955. IPW and Standardisation find $\overline{\overline{\beta_2}}$= 0.913 and 0.914, respectively. The first four methods underestimate the conditional causal effect, while the final two estimate accurately the marginal causal effect (IPW and standardisation).

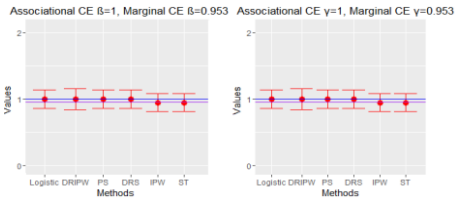


Figure 2. Estimates of β_1 (left) and β_2 (right), for the 'no unmeasured heterogeneity or confounding' scenario. Red: 95% confidence interval (CI) of estimates; blue: true conditional causal effect; purple: true marginal causal effect.

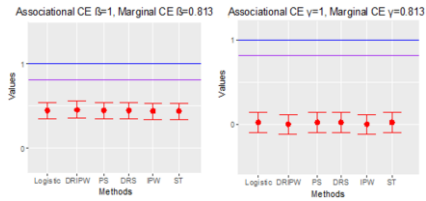


Figure 3. Estimates of β1 (left) and β2 (right), for the 'Unmeasured heterogeneity with confounding' scenario (β2= β4=1). Red: 95% confidence interval (CI) of estimates; blue: true conditional causal effect; purple: true marginal causal effect.

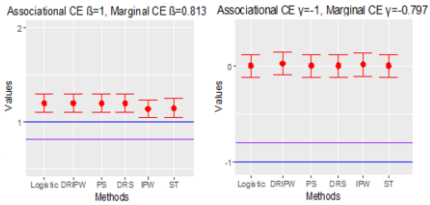


Figure 4. Estimates of β1 (left) and β2 (right), for the 'Unmeasured heterogeneity with confounding' scenario (β2= β4=-1). Red: 95% confidence interval (CI) of estimates; blue: true conditional causal effect; purple: true marginal causal effect.

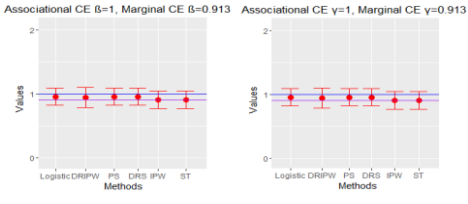


Figure 5. Estimates of β1 (left) and β2 (right), for the 'Unmeasured heterogeneity without confounding' scenario. Red: 95% confidence interval (CI) of estimates; blue: true conditional causal effect; purple: true marginal causal effect.

4. Discussion

This study found that unmeasured heterogeneity in the control treatments can lead to biased comparisons between the new intervention and the individual control treatments. This bias occurs when the heterogeneity in the use of control treatments is not fully measured (i.e., the individual treatments are used by different patient groups with varying risks for the outcome event of interest), whether or not it incorporates

unmeasured confounding (i.e. when the unmeasured variable also affects treatment allocation). We found that statistical techniques that are commonly used to correct for measured confounding did not overcome this bias. If unobserved heterogeneity and confounding exist, then treatment effects will be biased. We also found that this bias depends on the variance of the heterogeneity, its relation with the binary outcome, and its effect on treatment allocation (not shown here). In conclusion, unmeasured heterogeneity in the control treatment can lead to misleading conclusions about the comparative effectiveness of individual treatments.

Acknowledgements

The research leading to these results was conducted as part of Work Package 2 of the GetReal consortium. This paper only reflects the personal views of the stated authors. For further information please refer to http://www.imi-getreal.eu/. The work leading to these results has received support from the Innovative Medicines Initiative Joint Undertaking under grant agreement no. 115546, resources of which comprise financial contribution from the European Union Seventh Framework Programme (FP7/2007-2013) and EFPIA companies' in kind contribution.

References

[1] Segal, J. B., C. Weiss, and Varadhan R, Understanding heterogeneity of treatment effects in pragmatic trials with an example of a large, simple trial of a drug treatment for osteoporosis. White Paper. Center for Medical Technology and Policy (2012).
[2] Zwarenstein, M., Oxman, A. Why are so few randomized trials useful, and what can we do about it? Journal of Clinical Epidemiology, 59(11), 11251126, 2006
[3] Zwarenstein, M., Treweek, S. What kind of randomised trials do patients and clinicians need? Evidence-Based Medicine, 14(4), 101103, 2009
[4] Luce, B. R., Kramer, J. M., Goodman, S. N., Connor, J. T., Tunis, S., Whicher, D., Schwartz, J. S. Rethinking Randomized Clinical Trials for Comparative Effectiveness Research: The Need for Transformational Change, Medicine and Public Issues Annals of Internal Medicine, (June 2009), 206209
[5] van Staa TP, Dyson L, McCann G, Padmanabhan S, Belatri R, Goldacre B, Cassell J, Pirmohamed M, Torgerson D, Ronaldson S, Adamson J, Taweel A, Delaney B, Mahmood S, Baracaia S, Round T, Fox R, Hunter T, Gulliford M, Smeeth L. The opportunities and challenges of pragmatic point-of-care randomised trials using routinely collected electronic records: evaluations of two exemplar trials (2014), Health Technology Assessment, volume 18, issue 43, pp. 1 – 146
[6] Vickers, A. J. Clinical trials in crisis: Four simple methodologic fixes. Clinical Trials, 11(6), 615621, 2014
[7] Thorpe, K. E., Zwarenstein, M., Oxman, A. D., Treweek, S., Furberg, C. D., Altman, D. G., Tunis S. Bergel E., Harvey I., Magid D.J., Chalkidou, K. A pragmatic-explanatory continuum indicator summary (PRECIS): a tool to help trial designers. Journal of Clinical Epidemiology, 62(5), 464475, 2009
[8] Hernán MA, Robins JM (2016). Causal Inference. Boca Raton: Chapman & Hall/CRC, forthcoming
[9] Schmidt AF, Klungel OH, Groenwold RHH, Adjusting for Confounding in Early Post-launch Settings: Going Beyond Logistic Regression Models. Epidemiology. 2016 Jan; 27(1):133-42.
[10] Greenland, S. and Pearl, J., Adjustments and their consequences – collapsibility analysis using graphical models. International Statistical Review (2011), 79 401–426.

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-226

IT Infrastructure of an Oncological Trial Where Xenografts Inform Individual Second Line Treatment Decision

Doris LINDOERFER[a,1] and Ulrich MANSMANN[a]
[a] *Institute for Medical Information Processing, Biometry and Epidemiology (IBE), Ludwig-Maximilians-Universität München, Munich, Germany*

Abstract. Translational clinical research is often characterized by a unidirectional information flow from clinical to molecular data by using phenotypes to elucidate molecular disease processes. Here we present the RESIST study which uses xenograft information for individual treatment decisions after resistance to a specific anticancer treatment establishing a bidirectional information flow between patient and molecular biology. The paper discusses the specific challenges related to the IT infrastructure for such bidirectional translational projects and proposes solutions. A specific focus is the safeguarding genomic privacy.

Keywords. IT infrastructure, translational research, anonymization / pseudonymization

1. Introduction

Individualized (personalized, precision, stratified) medicine aims to fit the most effective treatment to an individual patient. The actual clinical standard is far from this goal. Often, expensive innovative cancer treatments fail because the biology of the individual cancer cell does not respond as expected. One strategy to optimize the individual treatment in cancer patients is to expose xenografts of the patient's tumor to different substances. Their response is supposed to inform the clinician on the patient's suitable treatment.

We study the scenario where patients develop resistance to a first line cancer treatment and need an appropriate second line approach which is derived from xenografts. A clinical trial and a biological sub-study is planned to assess the clinical effect of the xenograft strategy, to elucidate the cellular biology of resistance, and to find indicators for the prediction of effective second line therapies.

The information management within this setting is beyond the traditional IT infrastructures for clinical trials and we propose an extension for the above described setting.

The challenges met are the following: (data model and data integration) beside the clinical trial database we need a database for the complex biological data derived from

[1] Corresponding Author: Doris Lindoerfer, Dipl.-Inf.; Institute for Medical Information Processing, Biometry and Epidemiology (IBE), Ludwig-Maximilians Universität München, Marchioninistrasse 15, D-81377 Munich, Germany; Email: lindoerf@ibe.med.uni-muenchen.de

the xenografts (genomics, transcriptomics, proteomics, metabolomics - OMICS); Automatic data input from laboratory analysis machines; Involvement of several partners (oncologists, pathologists, biologist, and lab people); Aspects of data protection and access control, anonymization, pseudonymization; Identity management of patient, related xenografts, and related biomaterial; data input and management (query checks, completeness, timing control).

Figure 1 sketches the two parts of the translational colorectal cancer study "RESIST": First, a registry for an AVATARMODEL to which the patient has to agree to be registered and which documents the use of his/her tumor probes. The second part consists of a clinical trial: If a metastatic disease is diagnosed, the patient is treated first by chemotherapy and cetuximab, until resistance is diagnosed. For second line, the patient will be treated with the xenograft derived experimental treatment.

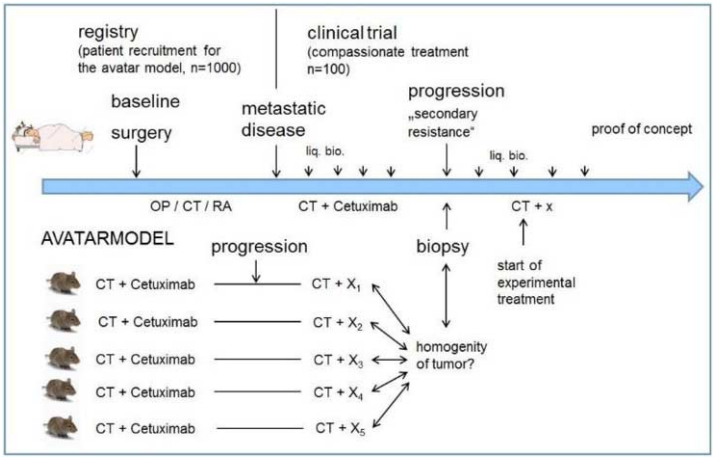


Figure 1. Schematic overview of the RESIST study

2. Methods

The RESIST study needs an IT infrastructure which combines a clinical database, an eCRF system, a data capturing system, and a database for information on the biomaterial (OMICS), a software tool which manages the collection and analysis of biomedical material as well as derived raw data. Anonymization / pseudonymization aspects need appropriate solutions. This is caused by the involvement of several medical centers, pathologies, labs, and biobanks. The following requirements need to be addressed: (1) Models for clinical and molecular data; (2) Data capturing tools (eCRF system); (3) Management of several identifiers (pids) related to the complex of patient – avatars – biomaterial (Special anonymization / pseudonymization mechanism); (4) Tools to support the project management (Information flow between the involved partners); (5) Interaction of clinical and OMICS database; (6) Access control; (7) Data management and quality control; (9) Data protection.

To meet these requirements we adopt our in-house eCRF system "dbform" [1] and combine it with recently developed tools. It runs on a platform providing a webserver and database management system (DBMS) environment under Linux, Apache [2], and

PostgreSQL [3], respectively. But other environments are possible. The major implementation programming language is Perl [4].

Our system provides role-based access control the rights can be granular assigned to people by study-groups, sites, screen forms and a small number of elementary rights. The server is behind a firewall and located in an access-controlled server room.

For the collection of the biomedical raw data open Biology Information System, Electronic Laboratory Notebook and Laboratory Information Management System (openBIS ELN-LIMS) [6] techniques will be used.

To enable the necessary information flow between the involved institutions and to guarantee an adequate pseudonymization / anonymization of the patient data we used techniques which are comparable to the for Germany developed generic data protection concept of the TMF e.V. [5] which need to be extended by instruments regarding Genomic Privacy [7], [8].

3. Results

Figure 2 shows the information flow in the RESIST study. Several clinical centers provide the patient data using the eCRF system. When a new patient is registered, the system creates several identification numbers: **pid** (the identifier for the clinical data); **tumor_no** (the identifier for the tumor material sent to pathology); **liqbio_no** (the identifier for the liquid biopsy (blood) sent to pathology)**; av_no** (the identifier for the connection to the biobank and other research institutions)

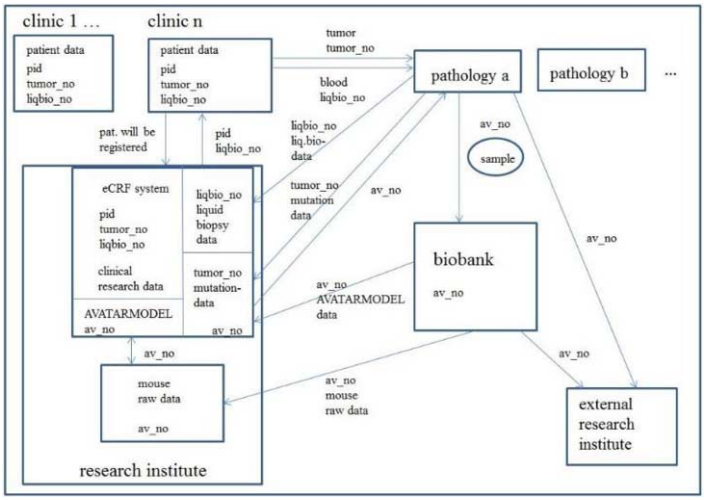


Figure 2. Information flow in a translational colorectal cancer research project.

The **pid** (Patient Identifier) is the identification number for the clinical data and is created when a patient is registered for the first time. This number is transferred into the patient's clinical health records to allow identification at a later stage. Clinicians have only access to the data of their own patients.

The clinic sends the tumor probe together with the **tumor_no** to the respective pathology where the tumor probe is analyzed and mutational data is inserted into the eCRF system. The pathology accesses the eCRF system with the **tumor_no** with a restricted access to see and insert selected data. The pathology extracts the **av_no** from the system and sends a sample of the tumor with the **av_no** to the biobank.

The biobank performs the AVATARMODEL research and captures the most important data of the AVATARMODEL (biological response, treatment schedule) with the **av_no** through the eCRF system into the database. The responsible persons from the biobank have only access to selected data in the database.

The biomedical avatar raw data from the biobank will be transferred regularly using a data-box to a second database to the research institute. The biomedical raw data is analyzed by an external research institute and entered into the system using the **av_no**.

Further, liquid biopsy (blood) will be sent regularly to the pathology and analyzed and liquid biopsy numbers (**liqbio_no**) will be created by the system.

The ethical review board required a completely anonymization of the data when the data capture is closed. Therefore, the **pid** will be removed from the database and in all files replaced by an independent indenture number. Thereby the connection between the research data in the database and the patient data in the clinics are capped and the research data are completely anonymized. For the identification and combination of the different data then the anonymous **av_no** will be used to analyze the data.

In order to provide genomic anonymity we process the molecular data using techniques proposed by Prasser et al. [7].

The overall RESIST data model is shown in Figure 3.

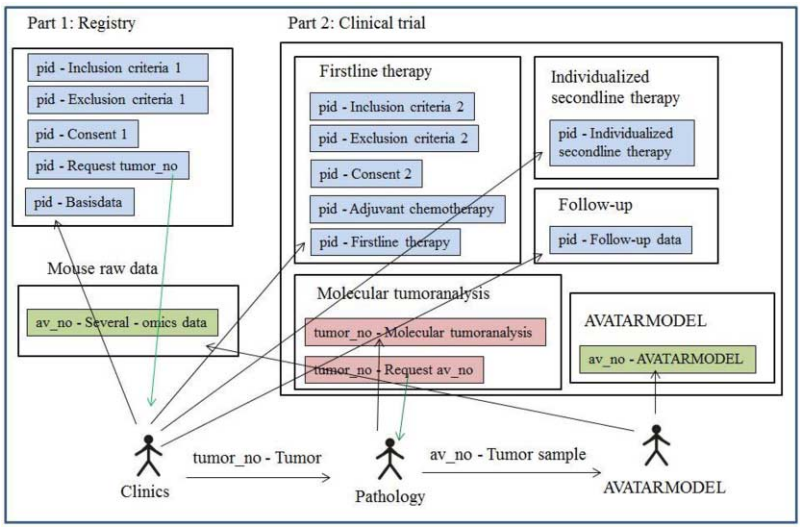


Figure 3. Simplified data model of the RESIST study.

The data related to the blue filled boxes are captured by the clinics, which requests the **tumor_no** and sends the tumor and **tumor_no** to the pathology. The pathology inserts molecular data with the **tumor_no** and requests the **av_no** and sends a tumor

sample with its **av_no** to the AVATARMODEL. AVATARMODEL group inserts data with the **av_no**. Compared to [9] we don't need to create safe havens, because the clinical study setting forces us to work in a safe haven. The data are protected through role-based data access and by different identifiers for different institutions.

The clinical data base will contain 1000 patients which is not challenging for the clinical data base. A total number of about 3000 xenografts are expected which creates a molecular data base which fits into the openBIS capacities. The RESIST study starts to recruit patients. The IT infrastructure was tested and validated by the joining clinicians, pathologists, and biologists.

4. Discussion

The implementation of biological systems into clinical decision making (like xenografts or AVATARMODELS as instrument for treatment selection) poses specific challenges for clinical trials. Here, the common separation between clinical database and databases for corresponding biological sub-studies are not appropriate as well as a unidirectional information flow (from the clinical database to the biological data to allow the analysis between phenotype and molecular processes). In contrast, the RESIST study implements a bidirectional setting: Molecular data derived from xenografts have to be transferred back in the clinical decision process.

The proposed IT infrastructure meets this bidirectional aspect by combining several open source tools: an in-house system, combined with the openBIS ELN-LIMS platform, and a management tool to organize the simultaneous processes at different institutions.

An important aspect is the anonymization / pseudonymization constraint regarding genomic privacy (implied by a series of different sources for the biomaterial, avatar, liquid biopsies, …) which needs handling safe haven besides the simple request of the ethical review board to remove patient related identifiers.

References

[1] Müller TH, Adelhard K: A web-based central diagnostic data repository. Stud Health Technol Inform. 2002, 90: 246-250.
[2] Apache HTTP SERVER PROJECT. https://httpd.apache.org/ (Accessed: Nov 07, 2016)
[3] PostgreSQL. Available: https://www.postgresql.org/ (Accessed: Nov 07, 2016)
[4] The Perl Programming Language. Available: https://www.perl.org/ (Accessed: Nov 07, 2016)
[5] TMF e.V. Available: http://www.tmf-ev.de/ (Accessed: Nov 07, 2016)
[6] Barillari C, Ottoz DS, Fuentes-Serna JM, Ramakrishnan C, Rinn B, Rudolf F. openBIS ELN-LIMS: an open-source database for academic laboratories. Bioinformatics. 2016 Feb 15;32(4):638-40. doi: 10.1093/bioinformatics/btv606.
[7] Prasser F, Kohlmayer F, Lautenschläger R, Kuhn KA. ARX – A Comprehensive Tool for Anonymizing Biomedical Data. AMIA Annual Symposium Proceedings 2014; 2014: 984 –993.
[8] Lautenschläger R, Kohlmayer F, Prasser F, Kuhn KA. A generic solution for web-based management of pseudonymized data. BMC Med Inform Decis Mak. 2015 Nov 30;15:100. doi: 10.1186/s12911-015-0222-y.
[9] Robertson D, Giunchiglia F, Pavis S, Turra E, Bella G, Elliot E, Morris A, Atkinson M, McAllister G, Manataki A, Papapanagiotou P, Parsons M. Healthcare data safe havens: towards a logical architecture and experiment automation. J Eng. 2016; doi: 10.1049/joe.2016.0170.

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-231

Medical and Healthcare Curriculum Exploratory Analysis

Martin KOMENDA[a,1], Matěj KAROLYI[a], Andrea POKORNÁ[a,b], Christos VAITSIS[c]
[a] *Institute of Biostatistics and Analyses, Faculty of Medicine, Masaryk University*
[b] *Department of Nursing, Faculty of Medicine, Masaryk University*
[c] *Department of Learning, Informatics Management and Ethics, Karolinska Institutet*

Abstract. In the recent years, medical and healthcare higher education institutions compile their curricula in different ways in order to cover all necessary topics and sections that the students will need to go through to success in their future clinical practice. A medical and healthcare curriculum consists of many descriptive parameters, which define statements of what, when, and how students will learn in the course of their studies. For the purpose of understanding a complicated medical and healthcare curriculum structure, we have developed a web-oriented platform for curriculum management covering in detail formal metadata specifications in accordance with the approved pedagogical background, namely outcome-based approach. Our platform provides a rich database that can be used for innovative detailed educational data analysis. In this contribution we would like to present how we used a proven process model as a way of increasing accuracy in solving individual analytical tasks with the available data. Moreover, we introduce an innovative approach on how to explore a dataset in accordance with the selected methodology. The achieved results from the selected analytical issues are presented here in clear visual interpretations in an attempt to visually describe the entire medical and healthcare curriculum.

Keywords. Exploratory data analysis, medical and healthcare curriculum, data mining, outcome-based education.

1. Introduction

Medical and healthcare education (MHE) is a domain which constantly needs to be evaluated and accordingly reshaped while it tries to incorporate to the extent possible the growing body of medical evidence. Higher education institutions aim to ensure both transparency and effectiveness as an educational system but also to create health professionals able to cope with healthcare trends and demands [1, 2]. This is a significant multifaceted task with underlying challenges and with a considerable level of complexity consisting of a number of components which need to be properly instrumented to effectively address these challenges. Although there is significant progress in using different techniques and methods such as different analytics approaches to leverage data successfully in other sectors, there is effort to transfer these techniques in higher education [3] but more process is required particularly for the complex world of MHE [4]. In previous studies, we have demonstrated how such

[1] Corresponding author, Kamenice 126/3, Brno, 625 00, Czech Republic; E-mail: komenda@iba.muni.cz.

educational data can be leveraged with the use of Visual Analytics in different cases. First in an effort to represent a medical and healthcare curriculum (MHC) from the perspective of relations between competencies and learning outcomes addressed in an entire medical programme [5]. Then, we used the same method to go deeper to analyze and investigate a specific course and its alignment for the same programme [6]. Finally, we evaluated the possibilities of all these approaches to support decision making [7] in an outcome-based MHE context. We have also demonstrated in another case and from another perspective how different analytical approaches, data analysis methods, and data representations can be used to collect the dispersed and multi-structured educational data to create a blueprint and fully or partially model MHC that would allow us to understand and reason for its different components and how they fit together in the big picture [8].

One of the most important aspects in data analysis is to be able to ask questions and get meaningful answers from the data. To achieve this, a deep understanding of the examined MHC as an entity is equally needed in order to give us the expertise and intuition needed to exploit the data we possess [9]. We have identified the following analytical issues, which are focused on how to systematically explore medical and healthcare study programmes from the data point of view. Specifically, the aim of this study is to investigate: (i) How can we effectively apply data analysis methodologies and techniques to identify hidden relations between MHC's components? (ii) How can we visually represent the identified hidden components and measurements in order to provide insights on the used pedagogical approaches behind developing and delivering learning activities, and analytically describe the entire MHC as an objective to support decision making?

2. Methods

In this section, we introduce the selected methodological model for data mining as well as data analysis and visualization background. In order to systematically describe our study programmes including all forms of teaching and learning activities, we developed a specialized curriculum management system call OPTIMED [10]. OPTIMED is an original platform for optimizing and harmonizing MHC, while supporting the outcome-based approach to education. Moreover, it provides huge volume of well-organized data stored in entity-relation database (detailed formal metadata specification down to the level of medical sections, disciplines, courses, learning units, and interconnections to the learning outcomes).

Many different data mining methodologies have been developed and are well established in order to systematically approach and solve tasks similar to MHC data analysis, modeling and deeper understanding. We have adopted CRISP-DM [11], which is often used for medical and healthcare as well as higher education data mining tasks [13–15]. This model completely fits to the curriculum exploration, which among other things fully supports the following crucial steps: (i) The understanding of the curriculum innovation objectives from the academic perspective, which are defined by research questions. (ii) The understanding of initial MHC data arrangement including the data quality problems identification. (iii) The evaluation and refinement of mined results by medical experts in a manner more effective to higher education institutions.

The six-stage-sequence CRISP-DM process helps us identify, analyze, and visualize hidden relations between MHC components as following: (i) The first step is

data extraction, where we accessed our PostgreSQL database to obtain raw data. (ii) The extracted dataset was then pre-processed, normalized and cleaned. Preprocessing phase is partially done by the database server and partially by server side application code written in PHP or R language. This produced a well-formed dataset with required data types and clear structure. (iii) At the next step we filtered the data to exclude entries irrelevant to the analysis purpose. What entries are irrelevant and therefore how the entire filtering approach is performed, is very much depended on the needs of the data reports recipients (e.g. curriculum designers, teachers), represented by the different visualizations. (iv) Following the data preparation we further analyzed the data with the use of available statistical software (SPSS, MS Excel, R). In this step, anomalies and discrepancies are frequently revealed in the dataset and if necessary the first three steps are repeated to refine the data and run the forth step again. This produces comprehensive static reports in a form of tables and graphs including legends and short data descriptions. (v) After evaluating the data reports, we automated the first three phases as described in previous subsection (phase iv) and thereafter this process enters a routine and executed daily. Thus, the generated reports are stored in a server where can be used in front-end modeling and visualization. (vi) The final step is modeling and visualization, where we produce two types of outputs: tables and graphs as static images, and web-based interactive graphs and tables with advanced filtering features. To plot the graphs, we used D3.js and NVD3 libraries and we followed the user-driven exploratory data analysis approach, which is based on additional filtering functionalities that enable the user to intuitively adjust visualizations and to explore the provided dataset by himself/herself.

3. Results

We created visual representations of the curriculum analysis, as a way to make the study programme more transparent and more easily understandable to students, teachers, faculty management, and potential employers. Graphs and tables show the connections and links between the various components of the curriculum. It enabled us to explore some general patterns within the study programme in relation to the promotion of generic skills. More than 25 various analytical reports have been prepared on the General medicine study programme. They show primarily basic overviews providing required summarizations in numbers. Below are two selected analytical reports from a pool of twenty-five analytical reports in total of explored hidden components and measurements in curriculum. We have integrated Bloom's educational activities in the curriculum conceptual data model including action verbs. Based on the presence of these meta-information in learning outcomes, we are able to categorize them in several levels and determine the requirements for students (Figure 1).

Figure 2 illustrates the ability to extract information about the annotation extent in a comprehensive manner. It is not humanly possible to go through the entire contents or to imagine how the sections, medical disciplines or courses are described by a set of plain text attributes, where a total length is not limited (e.g. meaning or annotation). We are able to determine their exact length in standard pages, compare individual curriculum components and their measurements, and provide the objective material to further discussion and support decision making.

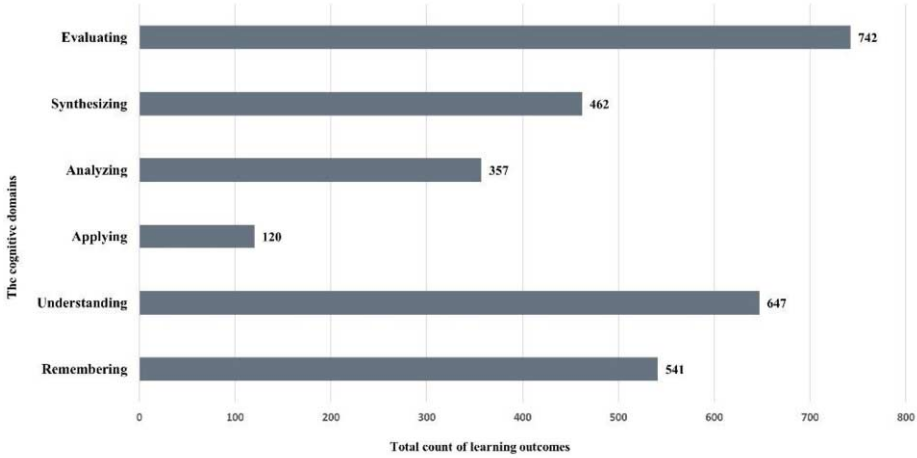


Figure 1. The content classification according to Bloom's taxonomy.

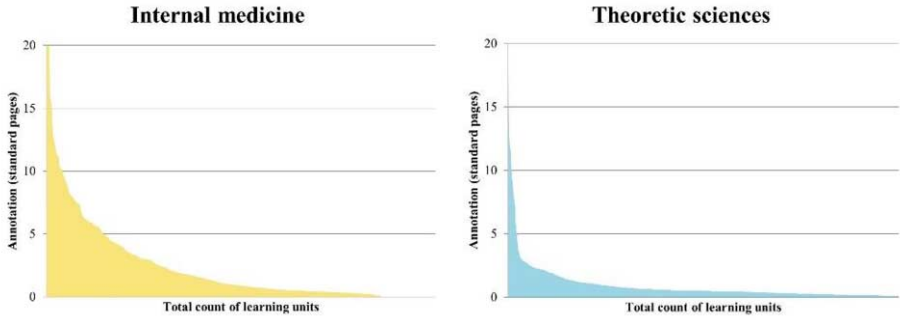


Figure 2. The annotation length of learning units according to medical sections.

4. Discussion

The exploratory analysis focused on curriculum evaluation should be always prepared very carefully in order to be of practical use for all stakeholders in the education process. We highlight with our results how analytical reports can be useful tools and how their usage in an appropriate way could contribute to the evaluation and improvement of the curriculum content in general. As we can see on Figure 1, there is wide range of Bloom´s taxonomy cognitive domains, which are covered in different learning units in analyzed curriculum. We could speak about success when evaluating the type of selected outcomes, because the ability to use Bloom's taxonomy by clinician-teachers was very low when we started to prepare the OPTIMED platform. Moreover, Figure 2 shows how we compare the annotation length or extent of learning units to different medical sections. Despite that the evaluation of the content is presumably not the main focus of this study, this information is very useful for guarantors of each medical section who could compare whether the learning units produced under their leading have the same basic components and to which amount.

We anticipate that our approach will support different stakeholders involved in MHE to build a deeper understanding of how MHE is currently conducted, use it for different education evaluation purposes and also to positively impact decisions concerning possible future reformations of MHE.

5. Conclusion

In general, the presented analytical reports show global overviews and identify whether the intended curriculum structure is actually being well-balanced. Moreover, we are able to demonstrate the overlapping or missing links among the different components of explored study programme. It helps mainly the staff by displaying these key elements of the curriculum and the relationships between them. Teachers and faculty management can be clear about their role in the big picture. Combining gained expert analytical knowledge with the CRISP-DM data mining process methodology can help reaching the core of a defined research questions. It can also advise the process of real data analysis and preparation, the features selection, the design and algorithms fine-tuning, and the evaluation and refinement of mined results in a manner more effective to higher education institutions.

References

[1] J. Frenk *et al.*, "Health professionals for a new century: transforming education to strengthen health systems in an interdependent world," *The lancet*, vol. 376, no. 9756, pp. 1923–1958, 2010.

[2] N. Cho, C. Gilchrist, G. Costain, and N. Rosenblum, "Incorporating evidence-based medicine in the undergraduate medical curriculum: early exposure to a journal club may be a viable solution," *Univ. Tor. Med. J.*, vol. 88, no. 3, pp. 154–155, 2011.

[3] C. Vaitsis, V. Hervatis, and N. Zary, "Introduction to Big Data in Education and Its Contribution to the Quality Improvement Processes," 2016.

[4] S. Mennin, "Self-organisation, integration and curriculum in the complex world of medical education," *Med. Educ.*, vol. 44, no. 1, pp. 20–30, 2010.

[5] C. Vaitsis, G. Nilsson, and N. Zary, "Big Data in Medical Informatics: Improving Education Through Visual Analytics," *Stud. Health Technol. Inform.*, vol. 205, p. 1163, 2014.

[6] C. Vaitsis, G. Nilsson, and N. Zary, "Visual analytics in healthcare education: exploring novel ways to analyze and represent big data in undergraduate medical education," *PeerJ*, vol. 2, p. e683, 2014.

[7] C. Vaitsis, G. Nilsson, and N. Zary, "Visual Analytics in Medical Education: Impacting Analytical Reasoning and Decision Making for Quality Improvement," *Stud. Health Technol. Inform.*, vol. 210, no. Digital Healthcare Empowering Europeans, pp. 95–99, 2015.

[8] M. Komenda, D. Schwarz, J. Švancara, C. Vaitsis, N. Zary, and L. Dušek, "Practical use of medical terminology in curriculum mapping," *Comput. Biol. Med.*, 2015.

[9] H. Cuesta, *Practical data analysis*. Packt Publishing Ltd, 2013.

[10] M. Komenda, D. Schwarz, C. Vaitsis, N. Zary, J. Štěrba, and L. Dušek, "OPTIMED Platform: Curriculum Harmonisation System for Medical and Healthcare Education," *Stud. Health Technol. Inform.*, vol. 210, pp. 511–515, 2015.

[11] P. Chapman, J. Clinton, T. Khabaza, T. Reinartz, and R. Wirth, "The CRISP-DM process model," *CRIP–DM Consort.*, vol. 310, 1999.

[12] N. Caetano, P. Cortez, and R. M. Laureano, "Using Data Mining for Prediction of Hospital Length of Stay: An Application of the CRISP-DM Methodology," in *Enterprise Information Systems*, Springer, 2014, pp. 149–166.

[13] D. Asamoah and R. Sharda, "Adapting CRISP-DM Process for Social Network Analytics: Application to Healthcare," 2015.

[14] C. Catley, K. Smith, C. McGregor, and M. Tracy, "Extending CRISP-DM to incorporate temporal data mining of multidimensional medical data streams: A neonatal intensive care unit case study," in *22nd IEEE International Symposium on Computer-Based Medical Systems, 2009*, 2009, pp. 1–5.

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-236

Using Electronic Health Records to Assess Depression and Cancer Comorbidities

Miguel A. MAYER[1], Alba GUTIERREZ-SACRISTAN, Angela LEIS, Santiago DE LA PEÑA, Ferran SANZ and Laura I. FURLONG
Research Programme on Biomedical Informatics (GRIB), IMIM, Universitat Pompeu Fabra, Barcelona, Spain

Abstract. Comorbid diseases are an important concern in oncology since they can affect the choice and effectiveness of treatment. What is particularly relevant is the fact that the diagnosis of depression in cancer patients has an important impact on the quality of life of these patients. Although there is no consensus about a specific relationship of depression with certain cancer types, some authors have proposed that depression constitutes a risk factor for cancer. The objective of this study is to identify the presence of comorbidities in a massive EHR system, between depression and the 10 most common cancers in women and men and to determine if there is a preferred temporal ordering in the co-occurrence of these diseases. All the cancers studied showed a significant co-occurrence with depression, more specifically, twice more frequent than what could be expected by chance. A preferred directionality was identified between some of the comorbid diseases, such as breast cancer followed by depression, and depression followed by either stomach cancer, colorectal cancer or lung cancer. Future work will address other potential factors that have an influence on the likelihood of suffering from depression in patients with cancer, such as drug therapies received, exposure to substance of abuse or other comorbidities.

Keywords. Comorbidity, Electronic Health Records, Depression, Cancer

1. Introduction

Comorbid diseases are an important concern in oncology since they can affect the choice and effectiveness of treatment [1,2]. Of relevance is the fact that the diagnosis of depression in cancer patients has an important impact on the quality of life of these patients, anticancer treatment choice and compliance and the health care costs as well [3,4]. More than one third of cancer patients suffer from psychiatric disorders, including depression, and the incidence rate is 2-3 times higher than in the general population [3]. Besides, depression is said to be the least recognised symptom in people with cancer [5] and in fact depression is associated with a decrease in survival following the diagnosis of cancer [1]. Although there is no consensus about a specific relationship of depression with certain cancer types [2], some authors have proposed that depression constitutes a risk factor for cancer [3]. In addition, it has been reported that chemotherapeutic treatment can result in mood disorders and impaired cognition [2,3].

[1] Corresponding author: Research Programme on Biomedical Informatics (GRIB), Dr. Aiguader 88, E-08003 Barcelona, Spain. E-mail: miguelangel.mayer@upf.edu.

According to the International Agency for Research on Cancer of the World Health Organization, the 10 most common cancers in men in Spain are, in order of frequency: prostate, lung, large bowel, bladder, pharynx (including oral cavity), stomach, kidney, liver, non-Hodgkin lymphomas and pancreas cancers. The 10 most common in women are: breast, large bowel, corpus uteri, lung, ovary, cervix uteri, malignant melanoma of skin, non-Hodgkin lymphomas, pancreas and stomach cancers [6]. In order to explore the comorbidity between cancer and depression, we performed a study on an EHR system from Spain. The objective of this study was first, to assess the presence of comorbidities between depression and the most prevalent cancers and second, for those cases in which a comorbidity is detected, to determine if there is a preferred temporal ordering in the diagnosis of both disorders. We report on the preliminary results obtained and point out future directions of this work.

2. Methodology

2.1. Electronic Health Records (EHR) System

The IMASIS information system is the Electronic Health Record (EHR) system of the Parc Salut Mar Barcelona Consortium, which is a complete healthcare services organization [7,8]. Currently, this information system includes and shares the clinical information of two general hospitals, one mental health care center, one social-healthcare center and five emergency room settings, which are offering specific and different services in the Barcelona city area (Spain). At present, IMASIS includes clinical information from patients who have used the services of this healthcare system since 1990 and from different settings such as admissions, outpatients, emergency room and major ambulatory surgery with a mean of 6.37 years (SD±6.82) of patient follow-up. The database contains hospital-based information on approximately 1.4 million patients and half of them have at least one diagnosis coded using The International Classification of Diseases ICD-9-CM [9]. IMASIS-2 is the relational model database containing anonymized patient information from IMASIS used for research purposes.

The study was performed on a corpus of 769,988 patient records from the IMASIS-2 database collected in the period 1990-2016. Based on the ICD9-CM codification all the patients diagnosed with depression (in accordance with the DSM-5 mental disorders manual) [10] and the 10 most common cancers in Spain were selected. Specifically, we selected breast, large bowel, corpus uteri, lung, ovary, cervix uteri cancers in women and prostate, lung, large bowel, bladder, pharynx (including oral cavity) and stomach cancers in men. The ICD9-MC codes used to select the patients are shown in table 1.

2.2. Analysis

The statistical analysis was performed using the R package comoRbidity [11]. This package contains several functionalities to assess the comorbidity among diseases, including the temporal directionality, sex ratio analysis and provides a variety of visualization tools. For each pair of diseases, the Relative Risk (RR) was calculated to estimate the degree of association between diseases (RR is estimated as the fraction between the number of patients diagnosed with both diseases and the random expectation based on the diseases prevalence) [12]. In addition, the Fisher exact test was performed to assess the null hypothesis of independence between the two diseases. The Benjamini

& Hochberg false discovery rate method [13] was applied to correct for multiple testing. From all the pairs of diseases, only those with corrected p-value lower than 0.01 and with a RR score greater than 1 were defined as comorbid diseases and further analyzed. We then assessed the temporal ordering of the association of two diseases (from disease A to disease B and from disease B to disease A) for the comorbidities identified in the analysis (directionality analysis). Specifically, the number of patients for whom diagnosis A follows diagnosis B, or vice versa, was calculated and an exact binomial test was subsequently used with a probability of success equal to 0.5. A preferred (significant) direction was assigned to those pairs of diagnoses in which the null hypothesis is rejected with $p < 0.05$.

Table 1. ICD9-CM codes used to select the diseases studied (depression and selected cancers)

Disease	ICD9-MC codes
Depression	295.7,295.71-295.75,296.2,296.20-296.26,296.3,296.30-296.36,296.4,296.40-296.46, 296.5,296.50-296.56,296.6,296.60-296.66,301.12,301.13,311,298.0,296.7, 296.8, 296.82, 296.89
Breast	174.9,174,174.0,174.1,174.2,174.3,174.4,174.6,174.8,174.9,233.0,238.3,239.3, V10.3
Bladder	156.0, 1561.1, 156.2, V10.51
Cervix	180,180.0,180.1,180.8,180.9,233.1,V10.41
Bowel	153.9,230.3,235.2,V10.05,V10.06
Lung	162.2,162.3,162.4,162.5,162.8,162.9,V10.11
Ovary	183.0, 233.9, 236.2, V10.43
Pharynx	149.0, 230.0, 235.1, 140.0-140.9, 145.9, V10.02
Prostate	185, 233.4, 236.5, V10.46
Stomach	151, 151.0-151.9, 209.23, V10.04
Uterus	179, 233.2,236.0,V10.42

3. Results

In IMASIS there are 367,041 women and 402,947 men with at least one diagnosis in ICD9-CM. 36,389 women and 26,742 men suffer at least from one of the studied diseases. The number of patients with each disease stratified by gender is shown in table 2. Depression is the most frequent disorder among the considered diseases, while the least frequent is corpus uteri cancer.

Table 2. Distribution of cases of the studied diseases stratified by gender.

Disease	Women	Men	Total
Depression	22,859	11,181	34,040
Breast	9,894	129	10,023
Bladder	665	2,681	3,346
Bowel	1,793	2,432	4,225
Cervix	1,014	-	1,014
Lung	1,074	5,208	6,282
Ovary	715	-	715
Pharynx	119	419	538
Prostate	-	6,181	6,181
Stomach	771	1,227	1,998
Uterus	622	-	622

As is shown in table 3 the Fisher exact test for each pair of diseases (depression with any cancer) indicates that there is a significant p-value and a Relative Risk (RR) > 1 in all

the cancers considered, with at least 30 patients in common in each pair of diseases, regardless of gender.

Table 3. Results of the analysis using the R package comoRbidity for association between depression and cancer (only significant p-values cases in the Fisher exact test for pairs of diseases are included in the table)

	Disease A	Disease B	Cases A & B	RR	Directionality
Women	Ut	Dp	131	3.383	No directionality
	Br	Dp	1165	1.893	From A to B
	Ce	Dp	164	2.600	No directionality
	St	Dp	117	2.434	From B to A
	Bo	Dp	378	3.392	From B to A
	Ov	Dp	119	3.018	No directionality
	Lu	Dp	208	3.108	From B to A
	Ph	Dp	31	4.184	No directionality
Men	Bl	Dp	267	3.590	No directionality
	Pr	Dp	406	2.367	From A to B
	Lu	Dp	376	2.600	No directionality
	Bo	Dp	232	3.749	No directionality
	St	Dp	94	2.920	No directionality
	Ph	Dp	46	3.943	No directionality

Bl: bladder; Bo: bowel; Br: breast; Ce: cervix; Dp: depression; Lu: lung; Ph: pharynx; Pr: prostate; St: stomach; Ut: uterus; RR: Relative Risk.

We then assessed the temporal ordering of the co-occurring diseases, to determine if there is a preferred temporal directionality by using a binomial test. The results are shown in table 3 under the directionality column. In some cases, as in stomach, lung, and bowel cancers in women, there is a preferred ordering in the disease onset where depression occurs previously to these cancers. In other cases, the diagnosis of depression occurs following the diagnosis of cancer such as breast cancer in women and prostate cancer in men. In this first phase of the study we have not analyzed differences of time span between the onset of the conditions considered that may offer additional information.

4. Discussion and Conclusions

The results here presented indicate that the comorbidity between depression and the most common cancers in both women and men is very common. Although it is expected to find some cases of depression in patients following the diagnosis of cancer, it is surprising that there are several associations in which depression precedes cancer. In this sense, some authors propose that depression can be a risk factor for cancer [3]. At present, the pathogenesis of this risk is not known and this work is a first step towards the analysis of comorbiditites between depression and cancer. Future work includes a detailed analysis of the specific diseases at code level (e.g. using each ICD9-CM codes) that may show significant comorbidities to gain further insights on the disease subtypes involved. In addition, the analysis of other factors that can have potential influence on suffering from depression in patients with cancer, such as the use of pharmaceutical treatments, substance of abuse or either the presence of other co-occurring disorders and comorbidities.

There are some limitations that should be mentioned. On the one hand, depression is an under diagnosed and complex condition, above all when it coincides with cancer, and it is necessary to analyze the existence of other factors such as treatments received, substance abuse exposure or diseases. On the other hand, the selection of patients is hospital-based and these results can be different from clinical information coming from primary care settings.

Acknowledgements

We received support from the IMI-JU under grant agreement no.115372 (EMIF), resources of which are composed of financial contribution from the EU-FP7 (FP7/2007-2013) and EFPIA companies in kind contribution, and from the EU H2020 Research & Innovation Programme 2014-2020 under grant agreements no. 634143 (MedBioinformatics: Creating medically-driven integrative bioinformatics applications focused on oncology, CNS disorders and their comorbidities).

References

[1] J.M. Geraci, C.P. Escalante, J.L. Freeman, J.S. Goodwin. Comorbid disease and Cancer: the need for more relevant conceptual models in Health Services Research. *Journal of Clinical Oncology* **23** (30) (2005), 7399–7404.
[2] R.G. Torta, V. Ieraci. Pharmacological management of depression in patients with cancer: practical considerations. Drugs **73** (11) (2013), 1131-45.
[3] M.B. Currier, CB Nemeroff. Depression as a risk factor for cancer: from pathophysiological advances to treatment implications. Annual Review Medicine **65** (2014), 203-221.
[4] F.M. Reich. Depression and cancer: recent data on clinical issues, research challenges and treatment approaches. *Current Opinion in Oncology* **20** (4) (2008), 353-359.
[5] M.L. Williams. Depression – the hidden symptom in advanced cancer. *Journal of the Royal Society of Medicine* **96** (12) (2003), 577-581.
[6] International Agency for Research on Cancer, World Health Organization. Available at: http://www.iarc.fr.
[7] M.A. Mayer, LI Furlong, P. Torre et al. Reuse of EHRs to support clinical research in a hospital of reference. *Studies in Health Technology and Informatics* **66** (2015), 856–890.
[8] J.J. Sancho, I. Planas, D. Domenech, M. Martin-Baranera, J. Palau, F. Sanz. IMASIS. A multicenter hospital information system – experience in Barcelona. *Studies in Health Technology and Informatics* **56** (1998), 35-42.
[9] International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM). Centers for Disease Control and Prevention. Available at: http://www.cdc.gov/nchs/icd/icd9cm.htm.
[10] American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders (Fith edition). DSM-5*. Arlington, United States of America, 2013.
[11] A. Gutiérrez-Sacristan. comoRbidity R package. Available at: https://bitbucket.org/ibi_group/comorbidity.
[12] C. Hidalgo, N. Blumm, A. Barabási, N. Christakis. A dynamic network approach for the study of human phenotype. *PLoS Computational Biology* **5** (4) (2009), e1000353.
[13] Y. Benjamini, D. Drai, G. Elmer, N. Kafkafi, I. Golani. Controlling the false discovery rate in behavior genetics research. *Behavioural Brain Research* **125** (2001), 279-284.

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-241

Improving Terminology Mapping in Clinical Text with Context-Sensitive Spelling Correction

Juliusz DZIADEK[1] , Aron HENRIKSSON and Martin DUNELD
Department of Computer and Systems Sciences, Stockholm University

Abstract. The mapping of unstructured clinical text to an ontology facilitates meaningful secondary use of health records but is non-trivial due to lexical variation and the abundance of misspellings in hurriedly produced notes. Here, we apply several spelling correction methods to Swedish medical text and evaluate their impact on SNOMED CT mapping; first in a controlled evaluation using medical literature text with induced errors, followed by a partial evaluation on clinical notes. It is shown that the best-performing method is context-sensitive, taking into account trigram frequencies and utilizing a corpus-based dictionary.

Keywords. spelling correction, terminology mapping, clinical text

1. Introduction

The increasing adoption of electronic health records (EHRs) provides access to vast amounts of digitized healthcare data, which is potentially very valuable. There are, however, challenges in analyzing EHR data, in particular when it comes in the form of unstructured text data, which is known to be noisy and contain a high degree of shorthand and misspellings [1,2]. To facilitate the secondary use of EHR data, clinical text needs to be mapped to medical ontologies like SNOMED CT [3], which exists in multiple languages and has become the de facto standard for the representation of clinical concepts. Mapping clinical text to ontologies allows us to tap into medical knowledge and to transform unstructured data into a form that can more readily be analyzed by computers.

Mapping clinical text to ontologies and standardized terminologies is, however, nontrivial, not least due to the abundance of misspellings. Systems that perform mapping in English clinical text exist, such as cTAKES [4]. However, existing methods tend to rely largely on dictionary look-up methods, which struggle with misspellings. The performance can conceivably be improved by detecting and correcting misspellings prior to the mapping process. While spelling correction of clinical text has received some attention for English [5,6], less has been done for other languages. Here, we evaluate the use of spelling correction methods to Swedish medical and clinical text, and evaluate their impact on SNOMED mapping.

[1] Corresponding author: jmail@op.pl

2. Methods & Materials

This paper explores various algorithms for spelling correction on two Swedish corpora: (1) a literature corpus, comprising edited journal articles, and (2) a clinical corpus, comprising notes from EHRs. While the motivating use case is to improve SNOMED CT mapping in clinical text, the medical corpus allows for the creation of a synthetic reference standard under the assumption that these edited texts do not contain spelling errors. The algorithms are thoroughly evaluated on the literature corpus for their ability to (1) detect and (2) correct misspellings, as well as to what degree the SNOMED CT mapping improves with the different spelling correction strategies. The algorithms are also evaluated on the clinical corpus by quantifying the number of additional SNOMED CT mappings that were made – through exact string matching – post spelling correction, in order to see how they fare on noisier input. Finally, a small-scale manual evaluation of a context-sensitive algorithm is carried out by a domain expert.

2.1. Algorithms

Correcting misspellings can be divided into two sub-tasks: (1) misspelling detection and (2) spelling correction. For misspelling detection, the Swedish spell checker Stava [7] and medical dictionaries are used. The list of candidate misspellings produced by Stava is filtered using general and domain-specific dictionaries: tokens that do not match any dictionary entry are treated as misspellings. For spelling correction, the baseline method is context-*insensitive* and based solely on Levenshtein distance. A number of context-*sensitive* methods, inspired by [5] and [6], are then evaluated and compared to the baseline method. Here, a method is defined as context-insensitive if it is deterministic w.r.t. to a token type, i.e., yield the same result irrespective of how often and where in the corpus it occurs. In contrast, the context-sensitive methods take into account not only the token type itself but also the context in which a particular token occurs and how frequent the token types are. Below is a description of the spelling correction algorithms. Two evaluations are carried out for each algorithm, employing a Levenshtein threshold of either 1 or 3 in the retrieval of replacement candidates.

Levenshtein Distance Candidates are retrieved from a dictionary and ranked according to Levenshtein distance to the misspelling, selecting the closest one. Candidates with the same Levenshtein distance are handled according to a source dictionary prioritization, whereby domain-specific dictionaries are preferred over general dictionaries. **Trigram Frequencies** Given the context of a misspelling in the form of a word trigram, where the misspelled word constitutes the middle word (or first/last if at beginning/end of sentence), the misspelled word within the trigram is replaced by any candidate with a Levenshtein distance $\leq$ t; the candidate with the highest trigram frequency in the corpus is selected. **Trigram Frequencies + Frequent Misspellings Filtering**: Like the previous algorithm, but with the difference that it employs a corpus-based dictionary which is used to filter out frequent candidates in the misspelling detection stage. **Trigram Frequency + Corpus-Based Dictionary**: Like Trigram Frequencies but employs a corpus-based dictionary which is used to filter out frequent candidates in the misspelling detection stage and, in contrast to the previous algorithm, also in the candidates retrieval stage. **Part-of-Speech Tagging + Frequent Misspellings Filtering**: When there are multiple candidates with the same Levenshtein

distance, those with the same part-of-speech are preferred. It also uses filtering of frequent tokens in the misspelling detection stage.

2.2. Experimental Setup

The medical literature corpus former comprising articles from the Journal of the Swedish Medical Association (1996-2005) [8], a subset of which (~1.3M tokens and ~0.2M types) is used. This corpus is treated as a reference standard. The algorithms are applied to a corrupted version of the corpus, in which spelling errors have been artificially introduced. Following [6], the probability of a token being misspelled is set to 15%; the misspellings are randomly introduced according to one of the four types of Damerau errors: insertion, replacement, transposition or deletion, as well as a compound error (i.e., white-space deletion). By comparing the corrected versions of the corrupted corpus to the original corpus, we are able to calculate precision, recall and F1-score for both misspelling detection and spelling correction. It moreover allows SNOMED mapping to be evaluated. The clinical corpus contains notes (~4.4M tokens and ~0.1M types) extracted from a database of Swedish EHRs1 Both corpora are tokenized using the Swedish spellchecker Stava [7] and part-of-speech tagged with Stagger [9]. Dictionaries were compiled from the Swedish versions of SNOMED CT [3], MeSH [10] and ICD-10 [11]; but also from NPL [12]: a Swedish registry of pharmaceutical products; Läkemedelsboken [13]: the Swedish Medical Products Agency's guidelines for pharmaceutical treatment; and SALDO [14]: a lexical resource for modern Swedish written language. The main evaluation criterion with the clinical corpus is to what extent more SNOMED mappings are possible post spelling correction. As this evaluation method ignores the notion of mapping accuracy, one of the context-sensitive algorithms is manually evaluated by a senior physician, effectively providing an estimation of its effectiveness.

3. Results

The results obtained for the two corpora are presented separately, beginning with the medical literature corpus. Spelling detection performance is high, particularly in terms of precision, with a score of 99.02%; recall is 81.72% and F1-score is 0.895. In contrast, the spelling correction module performs considerably worse (Table 1). Spelling correction precision varies between 48% and 71%, while recall oscillates between 14% and 26%.

Table 1. Spelling correction on the medical literature corpus

	Threshold=1			Threshold=3		
Algorithm	**Precision (%)**	**Recall (%)**	**F1-score**	**Precision (%)**	**Recall (%)**	**F1-score**
Levenshtein	69.22	14.74	0.243	58.23	**26.03**	**0.360**
Trigram	69.42	14.77	0.244	57.54	25.73	0.356
Trigram + Filtering	70.89	14.70	0.244	48.34	21.18	0.295
Trigram + Dictionary	**71.09**	14.74	0.244	58.08	25.44	0.354
POS + Filtering	69.47	14.40	0.238	53.54	23.46	0.326

The impact on SNOMED mapping on the literature corpus is shown in Figure 1, from which we can see that all algorithms lead to at least 7% additional token types being

mapped. Trigram + Filtering yields the biggest increase: 18.96%. Using a higher Levenshtein threshold invariably leads to better performance on both spelling correction and SNOMED mapping. In terms of mapping precision, however, Trigram + Dictionary performs best. In this case, using a lower Levenshtein threshold invariably yields better results. When comparing the former, the context-sensitive ones all outperform the context-insensitive baseline.

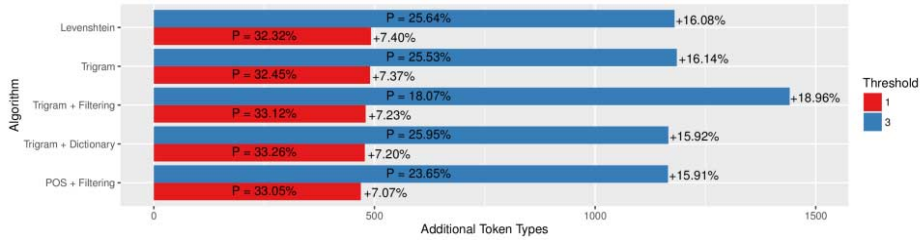


Figure 1. SNOMED mapping on medical literature.

The result of SNOMED mapping post spelling correction on the clinical corpus is shown in Figure 2, which reveals that all algorithms yield substantial mapping increases. As with the medical literature corpus, using a higher Levenshtein threshold leads to more mappings, with a context-sensitive algorithm resulting in the highest increase.

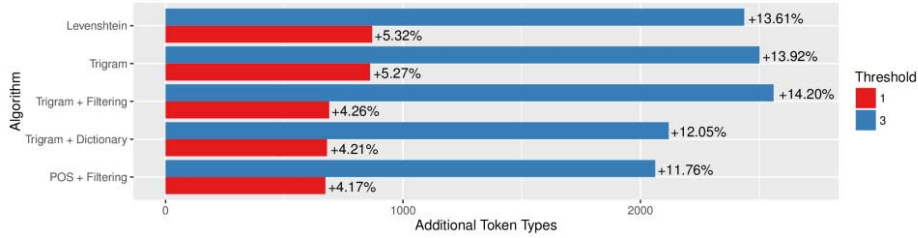


Figure 2. SNOMED mapping on clinical text.

A subset of the output (571 detected misspellings) from Trigram + Dictionary (with a threshold of 3) was then manually evaluated by a senior physician. Of these, 54.99% were categorized as spelled correctly; 38.88% were categorized as spelled incorrectly; while, for 6.13% of the tokens, the correctness could not readily be resolved. The sources of errors were spread across various token types: 68.30% were domain-specific words (including 12.78% drug names), while 17.16% were regular Swedish words. Moreover, 20.67% of the tokens were abbreviations, 21.72% inflections and 18.91% compounds. Of 222 corrected and confirmed misspellings, around 70% of the replacement candidates were marked as correct. Around 42% of the erroneous spelling correction candidates originated from the dictionary compiled from SNOMED CT.

4. Discussion

The manual evaluation revealed large differences in misspelling detection precision between the two corpora, indicating that this task is more challenging in the noisier

clinical text. It should be noted, however, that a general-purpose spell checker was employed for misspelling detection and that the only adaptation was in the form of domain-specific dictionaries. Filtering out candidates that were frequent in the corpus negatively affected performance on the medical literature corpus, probably because misspellings tend to recur. When misspellings had been correctly identified in the clinical corpus, however, replacement precision was moderate (70%). As expected, employing a lower Levenshtein threshold yields both higher spelling correction and mapping precision, while a higher threshold yields higher recall, as well as a larger number of SNOMED mappings. A context-sensitive algorithm, exploiting trigram frequency information and a corpus-based dictionary, obtained arguably the best overall results by yielding a large number of additional mappings with relatively high precision. This is encouraging, and it may be possible to obtain further improvements by taking into account additional context information: one option would be to leverage models of distributional semantics to this end. A more reliable evaluation of performance on clinical text would in the future require access to hand-annotated data.

References

[1] Helen Allvin, Elin Carlsson, Hercules Dalianis, Riitta Danielsson-Ojala, Vidas Daudaravi˘cius, Martin Hassel, Dimitrios Kokkinakis, Heljä Lundgren-Laine, Gunnar H Nilsson, Øystein Nytrø, et al. Characteristics of Finnish and Swedish intensive care nursing narratives: a comparative analysis to support the development of clinical language technologies. *Journal of Biomedical Semantics*, 2 (Suppl 3):S1, 2011.

[2] Kelly Smith, Beata Megyesi, Sumithra Velupillai, and Maria Kvist. Professional language in Swedish clinical text: Linguistic characterization and comparative studies. *Nordic Journal of Linguistics*, 37(02):297–323, 2014.

[3] IHTSDO. International Health Terminology Standards Development Organisation: SNOMED CT, 2015. Accessed: July 2, 2015.

[4] Guergana K Savova, James J Masanz, Philip V Ogren, Jiaping Zheng, Sunghwan Sohn, Karin C Kipper-Schuler, and Christopher G Chute. Mayo clinical text analysis and knowledge extraction system (ctakes): architecture, component evaluation and applications. *Journal of the American Medical Informatics Association*, 17(5):507–513, 2010.

[5] Jain Zheng Patrick, Sabbagh. Spelling Correction in Clinical Notes with Emphasis on First Suggestion Accuracy. In *Proc. of the International Conference on Language Resources and Evaluation*, 2010.

[6] Patrick Ruch, Robert Bauda, and Antoine Geissbuhler. Using lexical disambiguation and named-entity recognition to improve spelling correction in the electronic patient record. *Artificial Intelligence in Medicine*, 29(1-2):169–184, 2003.

[7] Kann Hollman. En metod för svensk rättstavning baserad på bloomfilter (In Swedish), 1993. Accessed: June 19, 2015.

[8] Dimitrios Kokkinakis. The Journal of the Swedish Medical Association – a Corpus Resource for Biomedical Text Mining in Swedish. In *Proceedings of the Third Workshop on Building and Evaluating Resources for Biomedical Text Mining (BioTxtM)*, 2012.

[9] Robert Östling. Stagger: An open-source part of speech tagger for Swedish. *Northern European Journal of Language Technology* (NEJLT), 3:1–18, 2013.

[10] NLM. U.S. National Library of Medicine: MeSH (Medical Subject Headings). http://www.ncbi.nlm.nih.gov/mesh, 2015. Accessed: July 2, 2015.

[11] WHO. World Health Organization: International Classification of Diseases (ICD), 2015. Accessed: July 2, 2015.

[12] Medical Products Agency. NPL National Repository for Medicinal Products, instructions for reviewing and verifying details in the NPL, 2011. Accessed: June 19, 2015.

[13] Läkemedelsverket. Läkemedelsboken (In Swedish), 2011. Accessed: June 19, 2015.

[14] Lars Borin, Markus Forsberg, and Lennart Lnngren. The hunting of the blark - saldo, a freely available lexical database for Swedish language technology. In *Resourceful language technology. Festschrift in honor of Anna Sågvall Hein*, pages 21–32. Uppsala University, Uppsala, 2008.

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-246

Medical Text Classification Using Convolutional Neural Networks

Mark HUGHES[a], Irene LI[a,b,1], Spyros KOTOULAS[a] and Toyotaro SUZUMURA[b,c]
[a] *IBM Research Lab, Ireland*
[b] *Japan Science and Technology Agency, Tokyo, Japan*
[c] *IBM TJ Watson Research Center, New York, USA*

Abstract. We present an approach to automatically classify clinical text at a sentence level. We are using deep convolutional neural networks to represent complex features. We train the network on a dataset providing a broad categorization of health information. Through a detailed evaluation, we demonstrate that our method outperforms several approaches widely used in natural language processing tasks by about 15%.

Keywords. Clinical text, semantic clinical classification, sentence classification, convolutional neural network

1. Introduction

Notes are key means of recording information about the health. Health professionals spend a lot of time scanning through notes with a view on identifying key problems and getting an overall impression of the status of the person. Particularly for complex cases that lead to information overload, delays or missing information [9]. With the notable exception of the works [1], current approaches rely on dictionaries to represent meaning, but introduces limitations at the representation and modeling levels, especially when representing social determinants of health. For example, the sentence "the patient lives with their mother, who is not able to leave her home" is important in a care management setting. It is hard to model using existing tools: first, the fragment refers to the patient's mother, rather than the patient; second, it indicates social exclusion without using any word that is associated with social exclusion by itself.

The goal behind this work is to apply machine learning approaches to build models that allow an automatically generated context-based and rich representation of health-related information. Convolutional neural networks (CNNs) have dramatically improved the approaches to many active research problems. One of the key differentiators between CNNs and traditional machine learning approaches is the ability for CNNs to learn complex feature representations.

We apply a CNN-based approach to categorization of text fragments, at a sentence level, based on the emergent semantics extracted from a corpus of medical text. We compare our approach with three other methods: Sentence Embeddings, Mean Word

[1] Corresponding author, Centre for Innovation, 7 Hanover Quay, Grand Canal Dock, Dublin, Ireland; Email: irenelizihui@gmail.com

Embeddings and Word Embeddings with BOW (bag-of-word). Our results indicate that the CNN-based approach is outperforming the other approaches by at least 15% in terms of accuracy in the task of classification.

2. Related Works

Classification for health-related text is considered a special case of text classification. Machine learning algorithms in Natural Language Processing (NLP) and have been successfully applied: e.g. Support Vector Machines and Latent Dirichlet Allocation have been used for some tasks like classification on patient record notes [1] or other documents in diseases like diabetes showing satisfying results [2,3]. The state-of-the-art models on document classification methods are designed for neural networks. Mikolov *et al*. [7] introduce an approach for learning word vector representations, Word2vec, which is simple and efficient. For neural embeddings, Le *et al*. [4] introduce the distributed representations of paragraphs, the Doc2vec, capturing the semantics in dense vectors. Other studies on CNNs for learning high level features have also shown competitive results. Other works by Kalchbrenner *et al*. [5] develop the Dynamic Convolutional Neural Networks for modelling sentences. This work is the first approach using such technology to do sentence-level classification of medical text.

3. Methods

We describe our method by means of a case study, where we have used Word2vec for a large corpus of text and a smaller corpus of pre-categorized text to train our sentence-level classifier.

We have used two datasets, procured from the medical domain. Our approach makes use of the Word2vec algorithm. It has been shown that performance can be improved by training Word2vec models using domain specific data. Therefore, we have procured a dataset from PubMed[1] for training our Word2vec models. To train our Word2vec models, we used a collection of 15k clinical research papers representing a wide range of medical subjects. The Word2vec model described in this paper was trained using this PubMed collection.

For sentence level classification, it was necessary to gather training data that had been pre-classified by medical professionals. Merck Manual[2] dataset contains articles from various topics like Brain, Cancer, etc. Each of these articles is classified under a parent header representing a specific category of medical issues and conditions. In total our dataset consisted of 26 medical categories and 4000 sentences were chosen at random for each category extracted from the Merck articles to use as our training data and to ensure balance across all categories. Our validation dataset consisted of 1000 sentences from each of the categories.

We apply a CNN-based approach to automatically learn and classify sentences into one of the 26 categories in our evaluation dataset. Similar to the approach outlined by

[1] PubMed is an online medical publication repository and contains published medical research across a very wide spectrum of clinical subjects. https://www.ncbi.nlm.nih.gov/pubmed

[2] Merck Manual is an online and offline resource containing encyclopedic style articles describing a wide range of medical subjects (http://www.merckmanuals.com/)

Kim [8], we convert each sentence to a word level matrix where each row in the matrix is a sentence vector extracted from our Word2vec model. CNNs require input to have a static size and sentence lengths can vary greatly. Therefore, we chose a max word length of 50 allowable for a sentence which worked well. During the training phase, we applied a Word2vec hidden layer size of 100, thus giving our input feature a resolution of 100×50. If a sentence contained less than 50 tokens, a special stop word was repeatedly appended to the end of the sentence to meet the 50-word requirement. If a sentence contained over 50 words, only the first 50 were considered to be representative of that sentence.

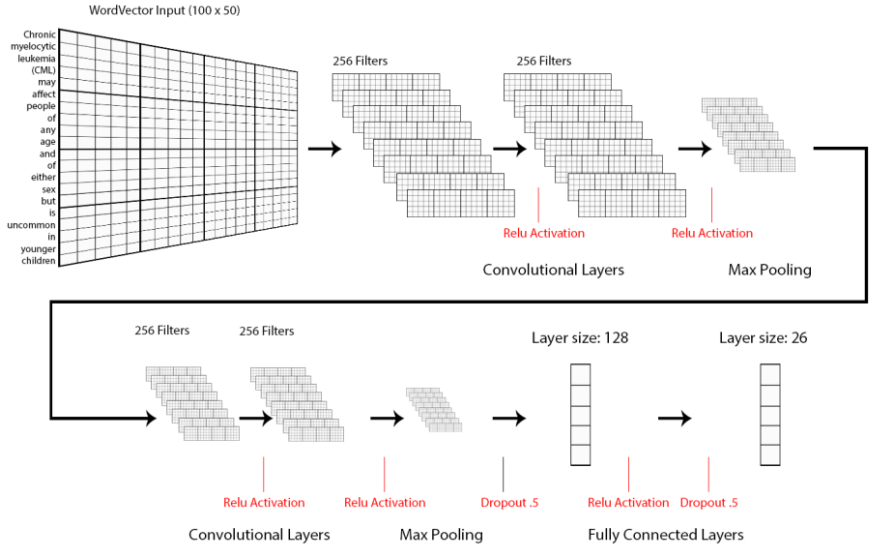


Figure 1. Outline of our CNN model structure.

During evaluation, we tested various CNN configurations. We applied a grid search to ascertain the optimal number of filters and filter sizes. We also experimented with multiple configurations of convolutional layers including 2, 4 and 6. From these experiments, our best performing CNN model consisted of a configuration of two sets of two convolutional layers with each pair followed by a max pooling layer. In this model, we used 256 convolutional filters with a filter size of 5 across all convolutional layers. After the second max pooling function we apply a dropout function to help preventing overfitting. In our model, we use a dropout rate of .5. We then append a fully connected layer with a length of 128 followed by a second dropout function. This is followed by a dense layer with a size of 26 to represent the number of classification classes with a Softmax function determining the output. A visual representation of this model can be found in Figure 1.

4. Evaluation

In this section, we evaluate our approach against a set of state-of-the-art methods. We have compared our model with the following methods: *Sentence Embeddings*, *Mean*

Word Embeddings and *Word Embeddings with BOW* (bag-of-word). During training word/sentence embeddings, the stop-words are kept and no stemming is adopted, since, in this way, we will keep the complete information. We keep the Doc2vec sentence embedding dimension to be 100, and the epoch to be 60 for all the experiments.

Sentence Embeddings (*LogR+Doc2vec*): Doc2vec, the *distributed memory* (PV-DM) model is firstly applied to train on the entire corpus. Once the model has been trained, each sentence in the test dataset can be inferred directly from the model. The second stage is to apply a Logistic Regression (LogR) classifier given the sentence embeddings inferred from the Doc2vec model.

Mean Word Embeddings (*ZeroMean/ElimMean+Word2vec*): For each sentence, we first took the embeddings of each word, and calculated a pair-wise mean as the sentence embedding. The lengths of sentences could be various - however, the dimension of the sentence embeddings is constant. When there is a word that is not in our vocabulary, we can fill with zeros or eliminate it (*ZeroMean+Word2vec* and *ElimMean+Word2vec* respectively). The mean word embeddings are the inputs to the above mentioned classifiers.

Word Embeddings with BOW Features (*BOW+LogR*): The third evaluation approach is the widely used Bag-of-Words histogram approach based on word vectors. As a part of this approach, Word2vec features are extracted from all entities within our dataset. We apply a k-means clustering algorithm with a value of 1000 for K to generate a feature vocabulary. Once this vocabulary is generated, a sentence is converted to a BOW histogram by assigning each word within the sentence to a vocabulary feature. Due to the short length of sentences in comparison to our vocabulary size, to avoid sparse vectors, we apply soft assignment when generating each BOW histogram. We use a value of 50 for K in the soft assignment phase with a value of 1/R appended to each histogram bin where R is the nearest neighbor ranking for the vocabulary feature associated with that bin up to a value of K.

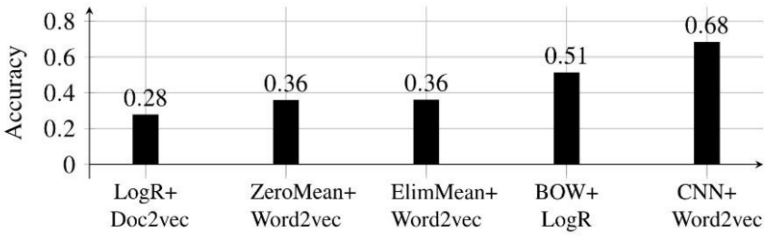


Figure 2. Classification Performance.

Figure 2 shows the accuracy (percentage of sentences classified correctly) of each method in our experiments. The first three methods shown in the figure give worse performance because the initially pre-trained embeddings are not providing good features for the classification. The bag-of-word method performs better - probably due to better feature extraction based on the pre-trained word embeddings. Our CNN-based approach has the highest accuracy by a wide margin. This could be explained by the fact that our deep learning approach has the ability to capture more complex features compared with the other shallow learning approaches.

5. Conclusions and Future Work

In this paper, we introduce a novel approach for sentence-level classification of medical documents. We show that it is possible to use CNNs to represent the semantics of clinical text enabling semantic classification at a sentence level. When compared with shallow learning methods, the multi-layer convolutional deep networks can generate more optimal features during the training phase to represent the semantics of the sentence being analyzed. Similar to computer vision methods, once these semantic representations are learned, they can also be used for many alternative tasks such as text comparison and retrieval tasks. With minimum effort, the approach could also be scaled up to generate representations at a paragraph or even document level.

In the future, we wish to implement our technique at a much larger scale and with a more fine-grained set of clinical classifications. We would expect that similar to the computer vision literature, our convolutional network approaches will provide better results when fed with larger datasets. Within the medical domain, we aim to test this hypothesis with a much larger collection of data collated from PubMed, relevant topics in Wikipedia, as well as medical books and journals. Alternatively, we also could leverage existing domain adaptation techniques [6,10] to transfer knowledge from other domains to the medical domain.

We are in the process of deploying a system using a similar technique with application to Care Management. Furthermore, we aim to experiment with the feature representations generated from a patient's clinical notes and apply them to generate a high level semantic representation of each patient. A patient could then be represented as a dense and highly discriminative feature vector that captures the medical conditions and treatments from their unstructured clinical notes, possibly combined with structured data.

References

[1] Cohen, R., Aviram, I., Elhadad, M., & Elhadad, N. (2014). Redundancy-aware topic modeling for patient record notes. PloS one, 9(2), e87555.

[2] Marafino, B. J., Davies, J. M., Bardach, N. S., Dean, M. L., Dudley, R. A., & Boscardin, J. (2014). N-gram support vector machines for scalable procedure and diagnosis classification, with applications to clinical free text data from the intensive care unit. Journal of the American Medical Informatics Association, 21(5), 871-875.

[3] Wang, L., Chu, F., & Xie, W. (2007). Accurate cancer classification using expressions of very few genes. IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB), 4(1), 40-53.

[4] Le, Q. V., & Mikolov, T. (2014, June). Distributed Representations of Sentences and Documents. In ICML (Vol. 14, pp. 1188-1196).

[5] Kalchbrenner, N., Grefenstette, E., & Blunsom, P. (2014). A convolutional neural network for modelling sentences. arXiv preprint arXiv:1404.2188.

[6] Sun, B., & Saenko, K. (2015). Subspace Distribution Alignment for Unsupervised Domain Adaptation. In *BMVC* (pp. 24-1).

[7] Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems* (pp. 3111-3119).

[8] Kim, Y. (2014). Convolutional neural networks for sentence classification. *arXiv preprint arXiv:1408.5882*.

[9] Pivovarov, R., & Elhadad, N. (2015). Automated methods for the summarization of electronic health records. *Journal of the American Medical Informatics Association*, *22*(5), 938-947.

[10] Sun, B., Feng, J., & Saenko, K. (2015). Return of frustratingly easy domain adaptation. *arXiv preprint arXiv:1511.05547*.

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-251

Acronym Disambiguation in Spanish Electronic Health Narratives Using Machine Learning Techniques

Ignacio RUBIO-LÓPEZ[a], Roberto COSTUMERO[a], Héctor AMBIT[a], Consuelo GONZALO-MARTÍN[a], Ernestina MENASALVAS[a], and Alejandro RODRÍGUEZ GONZÁLEZ[a,1]
[a] *Universidad Politécnica de Madrid, Centro de Tecnología Biomédica, Spain*

Abstract. Electronic Health Records (EHRs) are now being massively used in hospitals what has motivated current developments of new methods to process clinical narratives (unstructured data) making it possible to perform context-based searches. Current approaches to process the unstructured texts in EHRs are based in applying text mining or natural language processing (NLP) techniques over the data. In particular Named Entity Recognition (NER) is of paramount importance to retrieve specific biomedical concepts from the text providing the semantic type of the concept retrieved. However, it is very common that clinical notes contain lots of acronyms that cannot be identified by NER processes and even if they are identified, an acronym may correspond to several meanings, so disambiguation of the found term is needed. In this work we provide an approach to perform acronym disambiguation in Spanish EHR using machine learning techniques.

Keywords. Acronym disambiguation, natural language processing, machine learning, electronic health records, clinical notes

1. Introduction

Electronic Health Records (EHRs) and their use in medical institutions are becoming more and more popular and its adoption has been increasing during the last years [1]. An EHR contains a set of clinical notes with information regarding the history of the patients, including their signs or symptoms, previous diseases that have suffered or laboratory tests among other relevant information. The access to the unstructured information in EHRs (e.g. plain text with descriptions about a patient) is not easy. EHRs were created with several aims, including an easy interchange and retrieval of the information, however it still remains a difficult task to perform query and answering process in an accurate way.

Most of the works that have been applied to extract information from unstructured medical text are based on text-mining or natural language processing (NLP) techniques. In [2], a review of the kind of techniques applied in this context in clinical documents is presented. In the last years, several tools have appeared to perform automatic analysis of biomedical texts to identify concepts and relationships by means of NLP techniques.

[1] Corresponding Author: Alejandro Rodríguez González, Centro de Tecnología Biomédica, Campus de Montegancedo, Pozuelo de Alarcon, 28223, Madrid, Spain. Email: alejandro.rg@upm.es. ORCID: orcid.org/0000-0001-8801-4762.

Some successful approaches in this context of clinical settings are Apache cTAKES [3] and MetaMap [4]. Acronyms allow the physicians to speed-up the writing process minimizing the number of characters used, however the same combination of characters corresponds frequently to different concepts depending on the concrete context. Thus the disambiguation of the acronym is a challenge that remains an open problem. Another problem that arises when disambiguating acronyms is multilingualism: most of the NLP approaches developed over biomedical texts have been developed for English. In the case of Spanish language, only a few works have applied with varying successes [5-7].

In this work we focus on the automatic analysis of Electronic Health Records written in Spanish. Therefore, H2A architecture for NLP presented in [8] allows the analysis of free text gathered from EHRs by structuring relevant clinical information, is used. This architecture is based on a built-pipeline that is able to detect acronyms and associate them with all their corresponding meanings. Moreover, an approach to disambiguate the meaning of an acronym is needed. To fill the gaps, in this work we focus in the creation of an acronym disambiguation approach that uses machine learning techniques to learn the use and meaning of an acronym in each context.

2. Related Work

In this section we focus in identifying the main approaches to acronym and abbreviation disambiguation. This field, as well as NLP itself, is a vast field of research that has produced many results during the last years. For this reason, we focus on the main results produced in acronym disambiguation.

On one hand, some works are based on the automatic extraction and disambiguation of acronym-meaning pairs. These perform the automatic detection of the acronyms and their definitions allowing to automatically create a corpus of abbreviations and acronyms. Wren & Garner published a work based on heuristics to locate and identify acronym-definition pairs using subsets of Medline records obtaining more than 174,000 unique acronyms and 737,000 associated definitions [9]. Melton et al [10] have developed a method to improve sense inventories consisting in mapping long form expansions to concept in UMLS (Unified Medical Language System).

On the other hand, several works have been directly focused on the disambiguation of the terms found in the texts. Sergei Pakhomov developed one of the main papers published in the field. He developed a method in which a semi-supervised maximum entropy based approach for the normalization of abbreviations and acronyms in medical texts is applied [11]. In 2006, Joshi et al. [12] published a work for word sense disambiguation and acronym expansion based on kernel methods: one using knowledge derived from unlabeled text, a second one using semantic knowledge from ontologies and a third additive kernel consisted in the first two. The same authors performed a comparative study [13] of supervised acronym disambiguation methods in a corpus of clinical notes using three machine learning algorithms approaches which exceeds 90% of accuracy depending of the parameters. Most of the approaches developed have been based on knowledge-based techniques, being the machine learning one not very significant. Finally, none of the main efforts has been focused in the disambiguation of acronyms in Spanish documents.

3. Materials and Methods

Name Entity Recognition (NER) is the task of the NLP which is in charge of identifying concepts and their semantic types. Some of the tools previously mentioned like MetaMap or cTAKES make use of UMLS to retrieve these concepts. However, the amount of acronyms provided by UMLS for Spanish texts are not enough. For this reason the need to expand the list of acronyms in Spanish EHRs, led us to use the dictionary of medical acronyms available online at Sociedad Española de Documentación Médica[2].

Disambiguating medical acronyms is a complex task even for experts. As an example, "PCA" typically refers to "Posterior cerebral artery" or "Posterior communicating artery", both being arteries that are very close, belonging to circle of Willis and both can produce ischemia. An expert reading the term in the context of a transcranial duplex will understand that it is referred to the posterior cerebral artery. If misunderstood, surgery performed can differ and the patient could suffer consequences. Acronym disambiguation has been solved using machine learning as a binary classification task to determine if a definition is correct.

3.1. Corpus and dataset

The corpus used to create the models consists of a 150 clinical notes set of stroke patients and is known to be formed only with documents from the same area. The clinical notes with a higher number of acronyms have been selected. The reason to use a high-rank approach is to obtain a dataset that considers the maximum number of possible acronyms. The notes were cleaned up and manually annotated by the researchers. Once the clinical notes are annotated and curated, the datasets to train the models are generated. The training set consists in a CSV/ARFF[3] file where each line represents the acronym-definition pair found for each acronym together with the corresponding context features. Manual selection of features used for the training of the model (see table 1) is performed.

Table 1. Features used to train the model. Example sentence: "Fuerza proximal MII conservada y simétrica"

Feature	Description	Words in the example
pre-word, post-word	Words before and after the acronym.	proximal - conservada
acronym	The acronym.	MII
area	Area of report. Not used. Future feature.	N/A
definition	The definition obtained from SEDOM	Miembros inferiores
section	The section of the report where the acronym appears.	Exploración
PoS pre-word, PoS post-word	Part-of-Speech (PoS) of the word before and after the acronym.	ADJ – ADJ
label	True/False, if definition is correct in that environment.	True

3.2. Models creation

C4.5 decision trees were selected to create the models. WEKA [14] machine learning tool has been used for the creation and validation of such models. The validation of the models was performed using two approaches: a 10-fold cross validation and a validation using a test set of 30 clinical cases randomly selected (but not previously used in the training) and manually annotated from a total of 4,071 available clinical cases. In each of the datasets used there are a total of 7,148 acronym appearances with 480 unique

[2] http://www.sedom.es/
[3] https://weka.wikispaces.com/ARFF

acronyms. The acronym with a higher number of appearances is "U" (meaning "units" in most of the cases) with a total of 467 (6.53%).

Dataset 1 (D1) – baseline: Contains 6 features: pre-word, post-word, acronym, area, definition, and label. 9,627 instances exist after removing duplicates.

Dataset 2 (D2): Three more features have been introduced: PoS pre-word, PoS post-word, and the section in which the acronym is found within a document. We get 16,635 instances after removing duplicates. PoS helps to generalize the use of a meaning, as the same tags surround same meanings for a particular acronym, while others change the PoS of the surrounding tags. The section is an important factor, as the meaning of an acronym typically is the same within the same section.

4. Results

Two validation types have been performed: I) 10 fold cross-validation (CV); and II) Test-set from a 30 clinical notes subset randomly selected. Confusion matrices in Table 2 depict the values for performance of the algorithm using both validation methods.

Table 2. C4.5 Confusion Matrices

C4.5		Classification D1 – CV		Classification D2 – CV		Classification D1 – Test set		Classification D2 – Test set	
		False	True	False	True	False	True	False	True
Expert annotation	False	5935	249	10654	383	2082	315	2051	346
	True	404	3039	451	5147	49	490	45	494

From the confusion matrices in Table 2, precision, recall, and F1-score can be calculated, summarizing the results of CV. Table 3 shows the results using these metrics.

Table 3. C4.5 performance metrics

	C4.5 – CV			C4.5 Test-set		
	P	R	F1	P	R	F1
D1	93.2%	93.2%	93.2%	90.9%	87.6%	88.5%
D2	95.0%	95.0%	95.0%	90.7%	86.7%	87.7%

A 30 randomly selected EHRs set has been used for test-set, validating our approach using a realistic subset. The validation dataset contains 2,937 instances and 875 acronyms. Both models obtained a precision over 90%. These results are acceptable for this concrete problem and to be used and integrated into H2A's pipeline [8].

Confusion matrices (see table 2) show that results of test-set have around -8% F-score in comparison with cross-validation. The most reasonable explanation is that the documents used to generate the corpus were selected using the high-rank method. Therefore, a small bias could have been introduced. However, we consider that the random selection of the clinical notes does not ensure to correctly reduce the bias, though it enables a different test in a more "realistic" environment. D2 model using cross-validation performs better in terms of Precision, which is the goal of this experimentation. Nevertheless, in a "realistic" environment by performing the test-set, we can see that both models perform with similar patterns and the differences among them are not so relevant.

5. Conclusions and Future Work

In this paper we have presented a machine learning approach for medical acronyms disambiguation in Spanish EHRs. An important factor to gain trustworthiness from specialists has been the prioritization of models' precision versus their recall. In future works, the introduction of new features in the dataset that could be relevant, will be analyzed together with the performance of other machine learning techniques. It is worth noting that all the knowledge acquired for the acronym disambiguation could also be useful for concept disambiguation.

References

[1] Ofir Ben-Assuli. Electronic health records, adoption, quality of care, legal and privacy issues and their implementation in emergency departments. Health Policy, 119(3):287–297, 2015.
[2] Stéphane M Meystre, Guergana K Savova, Karin C Kipper-Schuler, John F Hurdle, et al. Extracting information from textual documents in the electronic health record: a review of recent research. Yearb Med Inform, 35:128–44, 2008.
[3] Guergana K Savova, James J Masanz, Philip V Ogren, Jiaping Zheng, Sunghwan Sohn, Karin C Kipper-Schuler, and Christopher G Chute. Mayo clinical text analysis and knowledge extraction system (ctakes): architecture, component evaluation and applications. Journal of the American Medical Informatics Association, 17(5):507–513, 2010.
[4] Alan R Aronson. Effective mapping of biomedical text to the umls metathesaurus: the metamap pro- gram. In Proceedings of the AMIA Symposium, page 17. American Medical Informatics Association, 2001.
[5] Roberto Costumero, Ángel García-Pedrero, Consuelo Gonzalo-Martín, Ernestina Menasalvas, and Socorro Millan. Text analysis and information extraction from spanish written documents. In International Conference on Brain Informatics and Health, pages 188–197. Springer, 2014.
[6] Maite Oronoz, Koldo Gojenola, Alicia Pérez, Arantza Díaz de Ilarraza, and Arantza Casillas. On the creation of a clinical gold standard corpus in spanish: Mining adverse drug reactions. Journal of biomedical informatics, 56:318–332, 2015.
[7] Elena Castro, Ana Iglesias, Paloma Martínez, and Leonardo Castano. Automatic identification of biomedical concepts in spanish-language unstructured clinical texts. In Proceedings of the 1st ACM International Health Informatics Symposium, pages 751–757. ACM, 2010.
[8] Ernestina Menasalvas, Alejandro Rodriguez-Gonzalez, Roberto Costumero, Hector Ambit, and Con- suelo Gonzalo. Clinical narrative analytics challenges. In International Joint Conference on Rough Sets, pages 23–32. Springer, 2016.
[9] Jonathan D Wren, Harold R Garner, et al. Heuristics for identification of acronym-definition patterns within text: towards an automated construction of comprehensive acronym-definition dictionaries. Methods of information in medicine, 41(5):426–434, 2002.
[10] Genevieve B Melton, SungRim Moon, Bridget McInnes, and Serguei Pakhomov. Automated identification of synonyms in biomedical acronym sense inventories. In Proceedings of the NAACL HLT 2010 Second Louhi Workshop on Text and Data Mining of Health Documents, pages 46–52. Association for Computational Linguistics, 2010.
[11] Serguei Pakhomov. Semi-supervised maximum entropy based approach to acronym and abbreviation normalization in medical texts. In Proceedings of the 40th annual meeting on association for computational linguistics, pages 160–167. Association for Computational Linguistics, 2002.
[12] Mahesh Joshi, Ted Pedersen, Richard Maclin, and Serguei Pakhomov. Kernel methods for word sense disambiguation and abbreviation expansion. In University of Minnesota. Citeseer, 2006.
[13] Mahesh Joshi, Serguei VS Pakhomov, Ted Pedersen, and Christopher G Chute. A comparative study of supervised learning as applied to acronym expansion in clinical reports. In AMIA, 2006.
[14] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H Witten. The weka data mining software: an update. ACM SIGKDD explorations newsletter, 11(1):10–18, 2009.

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-256

Automated Classification of Semi-Structured Pathology Reports into ICD-O Using SVM in Portuguese

Michel OLEYNIK[a,1], Diogo F. C. PATRÃO[b] and Marcelo FINGER[c]
[a] *Institute for Medical Informatics, Statistics and Documentation, Medical University of Graz, Austria*
[b] *International Center for Research, A.C.Camargo Cancer Center, Brazil*
[c] *Institute of Mathematics and Statistics, University of São Paulo, Brazil*

Abstract. Pathology reports are a main source of information regarding cancer diagnosis and are commonly written following semi-structured templates that include tumour localisation and behaviour. In this work, we evaluated the efficiency of support vector machines (SVMs) to classify pathology reports written in Portuguese into the International Classification of Diseases for Oncology (ICD-O), a biaxial classification of cancer topography and morphology. A partnership program with the Brazilian hospital A.C. Camargo Cancer Center provided anonymised pathology reports and structured data from 94,980 patients used for training and validation. We employed SVMs with tf-idf weighting scheme in a bag-of-words approach and report F_1 score of 0.82 for 18 sites and 0.73 for 49 morphology classes. With the largest dataset ever used in such a task, our work provides reliable estimates for the classification of pathology reports in Portuguese and agrees with a few similar studies published in the same kind of data in other languages.

Keywords. Natural language processing, pathology report, support vector machines

1. Introduction

Clinical reports are usually written in natural language due to its descriptive power and ease of communication among specialists. Processing data for knowledge discovery and statistical analysis requires information retrieval techniques, already established for newswire texts, but still in development in the medical subdomain. Some studies [1,2] explored mapping techniques to obtain structured information from clinical data, usually mixing sets of rules with machine learning. Although the results are promising, major efforts are required to build medical corpora and to adapt general language rules to the clinical domain.

Pathology reports are a main source of information regarding cancer diagnosis [3] and are commonly written following semi-structured templates that include tumour localisation and behaviour. Since structured data are mostly not available from the electronic health record (EHR) with sufficient accuracy and completeness, cancer

[1] Corresponding author: Michel Oleynik, Institute for Medical Informatics, Statistics and Documentation, Auenbruggerplatz 2, 8036 Graz, Austria; E-mail: michel.oleynik@stud.medunigraz.at.

registries play an important role as containers of manually reviewed clinical content at patient level, in order to report cancer statistics to authorities by using controlled vocabularies. In this process, health professionals often employ the International Classification of Diseases for Oncology (ICD-O) [3], a biaxial classification of cancer topography and morphology maintained by the World Health Organization (WHO).

Probably the first work to evaluate the automated classification of pathology reports into more than one class, Martinez and Li [4] explored a Naïve Bayes classifier with named entities as features to classify 217 reports written in English into 11 different sites. They report a micro-averaged F_1 score of 0.58. Jouhet et al. [5] employed Naïve Bayes classifiers and Support Vector Machines (SVMs) in 5,121 French free-text pathology reports to classify them into the two ICD-O axes using ngrams as features. They reported an F_1 score of 0.72 for 26 topographic sites and 0.85 for 18 morphology classes in the ICD-O code attribution task. Later, Kavuluru et al. [6] applied Naïve Bayes classifiers, SVMs and Logistic Regression to a dataset of 56,426 English pathology reports in order to classify them into 57 primary sites from the ICD-O. They compared the efficiency of unigrams, bigrams and named entities as features and reported an F_1 score of 0.90. More recently, Oleynik et al. [7] applied Naïve Bayes classifiers to a set of pathology reports written in Portuguese and obtained $F_1 = 0.75$ for the recognition of 16 topographies and $F_1 = 0.62$ for 49 morphologies.

In this paper, we applied SVM to a large dataset of pathology reports in Portuguese and assessed its efficiency. We report results in the two ICD-O axes, viz. topography and morphology. To our knowledge, there are no previous studies related to the same language, method and type of data.

2. Materials and Methods

2.1. Pathology Reports and Cancer Registry Corpora

A partnership program with the Brazilian hospital A.C. Camargo Cancer Center provided anonymised pathology reports and structured data from 94,980 patients used for training and validation. The documents were created during routine operation between 1996 and 2010, with their text structure following the institution's editorial guidelines.

In order to train a supervised machine learning classifier, we programmatically unified reports of the same patient and associated their content using the patient identifier to the data available in cancer registries. These registries include manually encoded information of the cancer topography, morphology and the metastatic status. In the next step, we discarded those with confirmed metastasis or multiple classifications and used the structured information in the cancer registries as the classifier target.

The resulting dataset maps patients into the two ICD-O axes, viz. topography with n = 18,905 patients (18 code groups) and morphology with n = 18,599 patients (49 code groups). Due to the nature of the problem, patients are not uniformly distributed in the groups, as can be seen on the data presented in Section 3.

2.2. Support Vector Machines

In a pre-processing step, we lowercased all tokens (extracted with the Java StringTokenizer class) and kept only the remaining 5,000 most frequent ones to speed up

processing and reduce model overfitting. We then applied a Support Vector Machine (SVM) over the vector space representation of the data (in a bag-of-words approach), with tf-idf weighting scheme[2] and a linear kernel. SVMs as such are known to produce good results in text classification [8].

A SVM is a discriminative and non-probabilistic classifier that tries to maximise the decision margin between two given classes [8]. The decision function, seen in Eq. (1), assigns either $+1$ or -1 to an input vector $\vec{x}$ based on the decision hyperplane normal vector $\vec{w}$ and an intercept term b.

$$f(\vec{x}) = sign(\vec{w}^T\vec{x} + b) \quad (1)$$

We used Weka 3.6.6 [9] for most of the steps and LibSVM 3.17 [10] to perform SVM calculations under a one-*versus*-all approach, common in multi-class classification tasks. A regularisation parameter $C = 2^{-7}$ was empirically determined following [11].

3. Results

Tables 1 and 2 show a breakdown of sample size (n), precision (P), recall (R) and F_1 score (F_1) obtained via 10-fold cross-validation for the ten topographies and morphologies with best results, respectively. The last row of each table shows the overall micro-averaged efficiency measures. Due to space limitations, full tables and confusion matrices are only available online at https://goo.gl/iG41Ok and https://goo.gl/cTrJfI.

Table 1. Top ten F_1 scores in the ICD-O topography attribution task.

Code Group	Description	n	P	R	F_1
C44	Skin	3,858	0.88	0.94	0.91
C50	Breast	3,668	0.89	0.91	0.90
C73-C75	Thyroid and other endocrine glands	1,329	0.92	0.87	0.90
C60-C63	Male genital organs	1,536	0.93	0.81	0.87
C64-C68	Lymph nodes	660	0.86	0.78	0.82
C51-C58	Female genital organs	1,574	0.85	0.77	0.81
C69-C72	Eye, brain and other parts of central nervous system	536	0.83	0.70	0.76
C00-C14	Lip, oral cavity and pharynx	903	0.80	0.71	0.75
C15-C26	Digestive organs	2,159	0.67	0.84	0.75
C77	Lymph nodes	590	0.68	0.80	0.74
	Overall	**18,905**	**0.82**	**0.82**	**0.82**

[2] Tf-idf stands for *term frequency–inverse document frequency*.

Table 2. Top ten F_1 scores in the ICD-O morphology attribution task.

Code Group	Description	n	P	R	F_1
959-972	Hodgkin and non-Hodgkin lymphomas	859	0.85	0.87	0.86
850-854	Ductal and lobular neoplasms	3,410	0.85	0.87	0.86
855	Acinar cell neoplasms	1,059	0.87	0.85	0.86
809-811	Basal cell neoplasms	1,704	0.80	0.89	0.84
872-879	Nevi and melanomas	1,473	0.87	0.81	0.84
906-909	Germ cell neoplasms	208	0.89	0.71	0.79
812-813	Transitional cell papillomas and carcinomas	384	0.81	0.74	0.78
938-948	Gliomas	237	0.82	0.71	0.76
858	Thymic epithelial neoplasms	17	1.00	0.59	0.74
868-871	Paragangliomas and glomus tumors	26	1.00	0.58	0.73
	Overall	**18,599**	**0.74**	**0.74**	**0.73**

4. Discussion

Results obtained in the topography attribution task ($F_1 = 0.82$) were better than the ones in the morphology attribution task ($F_1 = 0.73$). As expected, the evaluated efficiency is better at simpler tasks, where the number of target classes is lower (18 *versus* 49). The results might be dependent on the non-uniform distribution of classes in the dataset, with the best results reported on the most common classes. Additionally, the most common cancer in women (C50: Breast) and men (C60-C63: Male genital organs) achieved high precision rates (0.89 and 0.93, respectively).

Moreover, the analysis of the confusion matrix reveals that the most common mistake in the topography attribution task is the classification of C51-C58: Female genital organs as C15-C26: Digestive organs, which also accounts for its low precision (0.67) and is responsible for 5% (171/3380) of the incorrect classified patients. In the morphology axis, classification of 805-808: Squamous cell neoplasms as 814-838: Adenomas and adenocarcinomas is the most frequent source of error and accounts for 8% (381/4796) of the misclassifications.

Although a strict comparison is hard due to the lack of public data in the medical domain, SVM shows improvements over the Naïve Bayes approach of Martinez and Li [4] applied to the same kind of data. While they report an F_1 score of 0.58 with 11 sites, we achieved a higher F_1 score of 0.82 with seven more classes. Likewise, our study achieved results comparable to those reported by Jouhet et al. [5]. Even though we report a lower F_1 score of 0.74 (*versus* 0.85) with 31 more classes in the morphology code attribution task, we achieved a higher F_1 score of 0.82 (*versus* 0.72) over only eight less topography classes. Compared to the work of Kavuluru et al. [6], we obtained a lower F_1 score (0.82 *versus* 0.93) with four more target classes. Lastly, SVM performed much better than a prior work done in the same dataset with Naïve Bayes classifiers [7], with an F_1 score improvement of 0.07 and 0.11 in the topography and morphology groups, respectively. One common approach in these studies is to remove rare groups and therefore achieve better efficiency rates. In contrast, we report results in the standard set of 18 topography groups and 49 morphology groups, as defined by the WHO [3].

5. Conclusion

Although our classifier is in general agreement with other works reported in literature, we can see some limitations. A more precise analysis would have been done if we had classified a subset of the reports by a team of specialists without access to other patient data. Assessing Cohen's kappa factor among them would provide a smaller upper bound to the algorithm and perhaps reproduce high discordance rates described in literature. The unavailability of such a team for an extended period also grounds our automated process. Nonetheless, we could have tested other learning models known for providing better results, like neural networks. However, given its elevated complexity and high computational cost, we chose a simpler and more manageable approach.

Our study may improve recall rates in tasks such as cohort building for clinical trials, as it creates additional structured information over textual data. Moreover, it could be employed to ease manual classification of pathology reports via the generation of probability ordered code lists. The research showed that the automatic classification of pathology reports is not only feasible, but also achieves high efficiency rates comparable to those found in similar papers. We believe that our work provides a successful baseline for future research, not only for the classification of medical documents written in Portuguese, but also to be extended and applied to other domains.

Acknowledgments and Legal Aspects

We would like to thank Prof. Stefan Schulz for the paper revision. Our work is funded by the Brazilian National Research Council - CNPq (project number 206892/2014-4) and is approved by the committee on ethics on research of the A.C. Camargo Cancer Center (registered under number 1418/10).

References

[1] A. R. Aronson, "Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program", in *Proceedings of the AMIA Symposium*, p. 17, AMIA, 2001.

[2] C. Friedman, P. O. Alderson, J. H. M. Austin, J. J. Cimino, and S. B. Johnson, "A general natural-language text processor for clinical radiology", *JAMIA*, vol. 1, no. 2, p. 161, 1994.

[3] A. Fritz, C. Percy, A. Jack, K. Shanmugaratnam, L. Sobin, D. M. Parkin, and S. Whelan, eds., *International Classification of Diseases for Oncology*. WHO Press, third ed., 2000.

[4] D. Martinez and Y. Li, "Information extraction from pathology reports in a hospital setting", in *Proceedings of the 20th CIKM*, pp. 1877–1882, ACM, 2011.

[5] V. Jouhet, G. Defossez, A. Burgun, P. Le Beux, P. Levillain, P. Ingrand, V. Claveau, et al., "Automated classification of free-text pathology reports for registration of incident cases of cancer", *Methods of Information in Medicine*, vol. 51, no. 3, p. 242, 2012.

[6] R. Kavuluru, I. Hands, E. Durbin, and L. Witt, "Automatic extraction of ICD-O-3 primary sites from cancer pathology reports", in *Clinical Research Informatics AMIA symposium (forthcoming)*, 2013.

[7] M. Oleynik, M. Finger, and D. Patrão, "Automated classification of pathology reports", *StudHealth Technol Inform*, vol. 216, p. 1040, 2015.

[8] T. Joachims, "Text categorization with support vector machines: Learning with many relevant features", in *European conference on machine learning*, pp. 137–142, Springer, 1998.

[9] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The weka data mining software: an update", *ACM SIGKDD Explorations Newsletter*, vol. 11, no. 1, pp. 10–18, 2009.

[10] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines", *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 27:1–27:27, 2011.

[11] C.-W. Hsu, C.-C. Chang, C.-J. Lin, *et al.*, "A practical guide to support vector classification", 2003.

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-261

Informative Observation in Health Data: Association of Past Level and Trend with Time to Next Measurement

Matthew SPERRIN [a,1] , Emily PETHERICK [b] and Ellena BADRICK [a]
[a] *Farr Institute, Faculty of Biology, Medicine and Health, University of Manchester, Manchester Academic Health Science Centre*
[b] *School of Sport, Exercise and Health Sciences, Loughborough University*

Abstract. In routine health data, risk factors and biomarkers are typically measured irregularly in time, with the frequency of their measurement depending on a range of factors – for example, sicker patients are measured more often. This is termed *informative observation*. Failure to account for this in subsequent modelling can lead to bias. Here, we illustrate this issue using body mass index measurements taken on patients with type 2 diabetes in Salford, UK. We modelled the observation process (time to next measurement) as a recurrent event Cox model, and studied whether previous measurements in BMI, and trends in the BMI, were associated with changes in the frequency of measurement. Interestingly, we found that increasing BMI led to a lower propensity for future measurements. More broadly, this illustrates the need and opportunity to develop and apply models that account for, and exploit, informative observation.

Keywords. Informative observation, Longitudinal modelling, Observation processes.

1. Introduction

When conducting longitudinal statistical analysis with routinely collected health data, it is often assumed that the process that governs whether and when data are observed – the observation process – is *ignorable*. This statistically defined term means that we do not need to concern ourselves directly with the observation process, and it is not necessary to model the process explicitly. In real terms this translates to assuming that measurements of a risk factor or biomarker are regularly spaced (e.g. measured once per year), or that they are irregularly spaced but the spacing is not informative (conditional on measured covariates). There is an approximate correspondence with the related concepts in missing data of missing completely at random and missing at random.

Intuitively, however, observations are made according to an underlying process driven by the patient, the clinician, and the environment. Therefore, the timing of observations may be informative, over and above the actual values observed (again, this corresponds to missingness not at random). For example, a patient concerned about

[1] Corresponding author, Vaughan House, University of Manchester, M13 9GB;
E-mail: matthew.sperrin@manchester.ac.uk.

their health may engage with the health service more and hence have smaller gaps between measurements. A clinician concerned about the health of a patient may request to see them again sooner. The propensity for a subsequent observation of a risk factor may also depend on previous observed values, and trends in previous values of the risk factor. For example, if a biomarker is rapidly rising, the clinician may wish to measure it again within a short time period. This is termed *outcome-dependent follow-up*. Some methods to handle data subject to a non-ignorable observation process are available in the statistical literature [1], [2]; these are based on assuming a joint model for both the observation process and the outcome process. However, these have seen limited application to routinely collected health data. Moreover, existing approaches typically view the observation process as a 'nuisance' and not to be of scientific interest [3]; we hypothesise that the observation process can be exploited to gain additional information for inference.

In this paper we explored the properties of the observation process in the example of body mass index (BMI) measures for patients with type 2 diabetes (T2D), taken in primary care in Salford, UK. We hypothesised that BMI measurements would depend not only on patient demographics but also on previous measurements, and the current trend, of BMI – i.e. outcome-dependent follow-up.

2. Methods

We used anonymized primary care data from the Salford Integrated Record. Salford, UK, is a relatively deprived city in Greater Manchester, UK, with a population of approximately 300,000, served by a single hospital and 53 GP practices. Our study period was 1 April 2004 to 31 December 2012. The start date was chosen to align with the quality and outcomes framework (QOF), which is a scheme, started in 2004, in which GPs are incentivized to meet a range of indicators that promote patient care; one of these indicators is that T2D patients have a BMI measurement within the financial year. Individuals were considered 'at risk' for a BMI measurement during this period provided that they had received a T2D diagnosis and were alive. BMI measurements outside of this time range were excluded; however, they were used where appropriate as 'previous BMI' readings. An individual may have no BMI readings recorded at all, but still be included in the analysis (since they are still 'at risk' of a BMI measurement). If an explicit T2D diagnosis date was not available, the date of first prescription of anti-diabetic medication was used as a proxy for this. If neither of these were available the patient was removed from the analysis. Patients were also removed if no date of birth was available. Finally, patients who were younger than 35 or older than 85 at diagnosis date were also removed.

We built a statistical model that focused on the observation process (times at which BMI is observed) rather than the outcome itself (the BMI measurements). Specifically, we used a Cox proportional hazards model for recurrent events to model time to next BMI measurement. We used age as the timescale, and left truncated at the study start date or diabetes diagnosis date, whichever was later. The earliest of death and the study end date was considered a right censoring event. The model incorporated frailty terms to capture within-person correlation [4].

In our multivariable model, covariates underlying the observation process of primary interest were: the previous BMI measure; the difference between the previous measure and the one before, which represents a trend that would be observable by the

GP; and the difference between the current (yet to be taken) BMI measure and the previous one, which may reflect the patient's current perception of weight change. Time since diagnosis and calendar year were included as time-updated terms. We also included gender, and separate indicator variables for diagnosis of coronary heart disease (CHD), chronic obstructive pulmonary disease (COPD), asthma and cancer.

All analyses were carried out using R version 3.2.0 [5].

3. Results

A data exclusion flow chart, both at the patient level and individual BMI observation level, is given in Figure 1; the final dataset comprised 11,805 patients with a total of 133,425 BMI readings.

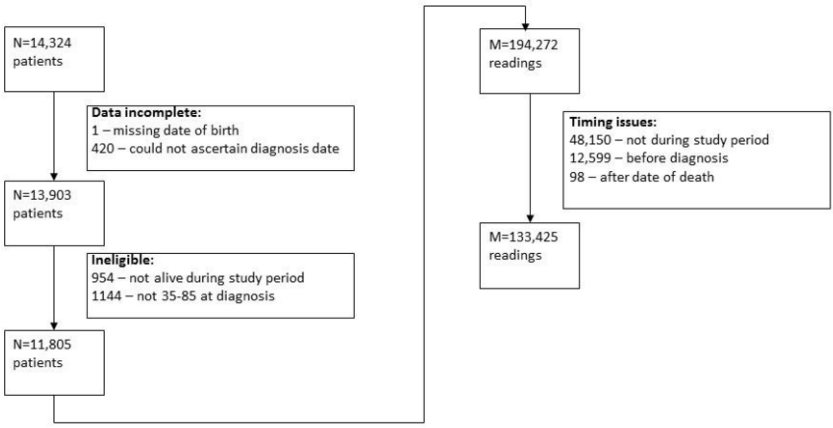


Figure 1. Data exclusion flowchart. Ineligible patients are excluded first, then ineligible readings are removed for eligible patients.

Baseline information is given in Table 1. The mean first BMI was 31.28 (in the obese category, which is as expected for T2D patients), and we observed a median of 9 BMI measurements per patient.

Table 1. Baseline information. (SD = standard deviation; IQR = interquartile range).

BMI at baseline (first reading in time period)	Mean = 31.28, SD = 6.41
Number of BMI measurements	Median = 9, Min=0, Max=112, IQR = 10
Year of birth	Median = 1944, Min = 1911, Max = 1976
Age at baseline	Mean = 62.02, SD = 11.74
Male	N = 6647 (56.31%)
CHD (ever)	N = 4183 (35.44%)
COPD (ever)	N = 1744 (14.77%)
Asthma (ever)	N = 2268 (19.21%)
Cancer (ever)	N = 1463 (12.39%)
Dead before 31/12/2012	N = 539 (4.57%)

Table 2 gives the proportion of patients for whom at least one BMI measure is made within a financial year (out of all patients who are alive and have a T2D

diagnosis for the entire financial year). This proportion increases steadily from 0.739 to 0.825 across the study period.

Table 2. Proportion of patients for whom at least one BMI measure is made within a QOF year (out of all patients who are alive and have a T2D diagnosis for the entire QOF year).

Year	Proportion with BMI reading	Number eligible
2004-5	0.739	6345
2005-6	0.752	6962
2006-7	0.780	7659
2007-8	0.806	8292
2008-9	0.815	8958
2009-10	0.826	9383
2010-11	0.834	9776
2011-12	0.825	10039

Results of the recurrent event proportional hazards model are given in Table 3, with hazard ratios given per unit BMI or per year as appropriate. We see that propensity (or hazard) to measure BMI was increased if the previous BMI reading was higher, but decreased if there was an observed or perceived upward trend in BMI. The difference between the two previous BMI readings had a larger effect on the hazard than the difference between the current and previous readings. CHD, COPD and asthma patients all had a higher propensity/hazard to be measured, while for cancer there was no significant difference in the hazard. There was no evidence of a violation of the proportional hazards assumption (P=0.957 in global test, and no individual covariates had significant relationship between Schoenfeld residuals and time).

Table 3. Hazard ratios from Cox recurrent event model.

Variable	HR (95% CI)
Previous BMI	1.010 (1.008,1.012)
Difference between previous reading and reading before	0.979 (0.976,0.983)
Difference between current and previous reading	0.985 (0.982,0.988)
Male (reference: female)	0.996 (0.973,1.021)
Calendar year	0.954 (0.951,0.957)
Time since diagnosis	1.027 (1.025,1.030)
CHD presence	1.026 (1.001,1.052)
COPD presence	1.127 (1.089,1.165)
Asthma presence	1.057 (1.024,1.090)
Cancer presence	0.983 (0.948,1.018)

4. Discussion

Contrary to our prior expectation, an increasing trend in BMI lowered the propensity for a repeat measurement of BMI. This is surprising and potentially concerning, and needs to be understood clinically. It was reassuring to find no evidence of a gender difference.

The findings show that the propensity to measure BMI depends on previous measurements and trends. This is likely to hold for other measures such as blood pressure and cholesterol. Appropriate statistical modelling techniques need to be used to account for this outcome dependent, non-ignorable structure in the observation

process, to prevent biased inference. One such approach is to build joint models for the observation process and outcome process (e.g. [6]). Mixed effect models can also be applied with limited bias for estimation of fixed effects [7]; however estimation of random effects can be badly biased [3].

Rather than viewing informative observation as a nuisance, we suggest that the presence of observations can be used for prediction. For example, we have shown that engagement with smart weighing scales (i.e. presence of weight measurements) is an independent predictor for weight loss [8].

The main strength of the study is that we use sophisticated modelling techniques to understand the observation process, which is typically ignored in the literature. A limitation is that we only considered BMI, although we expect that the findings will generalise to other clinical risk factors and biomarkers that may be measured irregularly over a patient's life course. Modelling limitations include that we have used ever/never terms for other diseases and smoking status – a time dependent approach could also have been considered. We could in theory also have incorporated other variables like blood pressure into the model; however they are themselves subject to irregular and potentially outcome-dependent follow-up. Moreover, a number of other variables were excluded that could explain changes in BMI – particularly T2D treatments such as metformin. We took a pragmatic approach to variable inclusion in this paper as we sought only to demonstrate the concept.

This study has shown in a real example that the observation process of a clinical risk factor may depend on previous measurements. It may also depend on other factors, measured or unmeasured. This is an area that brings challenge and opportunity: the challenge to produce models that are not biased by the presence of informative observation, and the opportunity to use the observation process itself in prediction.

References

[1] H. Lin, D. O. Scharfstein, and R. A. Rosenheck, "Analysis of longitudinal data with irregular, outcome-dependent follow-up," J. R. Stat. Soc. Ser. B (Statistical Methodol., vol. 66, no. 3, pp. 791–813, 2004.

[2] J. Sun, D.-H. Park, L. Sun, and X. Zhao, "Semiparametric Regression Analysis of Longitudinal Data With Informative Observation Times," J. Am. Stat. Assoc., vol. 100, no. 471, pp. 882–889, Sep. 2005.

[3] C. E. McCulloch, J. M. Neuhaus, and R. L. Olin, "Biased and unbiased estimation in longitudinal studies with informative visit processes," Biometrics, vol. 72, no. 4, pp. 1315–1324, Dec. 2016.

[4] C. McGilchrist and C. Aisbett, "Regression with frailty in survival analysis," Biometrics, 1991.

[5] R. Team, "R Development Core Team," R A Lang. Environ. Stat. Comput., 2013.

[6] J. Sun, L. Sun, and D. Liu, "Regression Analysis of Longitudinal Data in the Presence of Informative Observation and Censoring Times," Journal of the American Statistical Association, vol. 102. pp. 1397–1406, 2007.

[7] S. R. Lipsitz, G. M. Fitzmaurice, J. G. Ibrahim, R. Gelber, and S. Lipshultz, "Parameter Estimation in Longitudinal Studies with Outcome-Dependent Follow-Up," Biometrics, vol. 58, no. 3, pp. 621–630, 2002.

[8] M. Sperrin, H. Rushton, W. G. Dixon, A. Normand, J. Villard, A. Chieh, and I. Buchan, "Who Self-Weighs and What Do They Gain From It? A Retrospective Comparison Between Smart Scale Users and the General Population in England," J. Med. Internet Res., vol. 18, no. 1, p. e17, Jan. 2016.

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-266

Multivariate and Longitudinal Health System Indicators

Guido Antonio POWELL[a1] , Yu T LUO[a,b], Aman VERMA[a], David A STEPHENS[b], and David L BUCKERIDGE[a]
[a] *Surveillance Lab, McGill Clinical and Health Informatics,*
[b] *Mathematics and Statistics, McGill University, Montreal, Quebec, Canada*

Abstract. Within population health information systems, indicators are commonly presented as independent, cross-sectional measures, neglecting the multivariate, longitudinal nature of disease progression, health care use, and profiles of performance. We use administrative claims data of Montreal, Canada to identify patterns across indicators and over time in chronic obstructive pulmonary disease patients. We first cluster regions based on four health service indicators. Our second approach discovers individual-level trajectories based on a hidden Markov model using the same four indicators. Both approaches offer additional insights by facilitating the discovery and interpretation of indicators, such as a dual interpretation of low use of general practitioner services. These approaches to the analysis and visualization of health indicators can provide a foundation for information displays that will help decision makers identify areas of concern, predict future disease burden, and implement appropriate policies.

Keywords. Health Information Systems, Chronic Obstructive Pulmonary Disease, Health Indicators

1. Introduction

Support for decision making in healthcare system improvement requires informative indicators of healthcare quality and performance. These indicators can help accomplish several objectives, such as identifying priorities for resource allocation based on variance across regions [1] and identifying which regions are outliers given an expected distribution [2]. Analyses of such indicators are often performed independently and cross-sectionally, omitting much of the context available from multivariate and longitudinal approaches. An indicator of health service use illustrates how indicators can benefit by expanding analyses on several dimensions. Consider, for example, the proportion of patients visiting a general practitioner (GP) in the past year. Such an indicator is typically presented in a stratified manner to facilitate comparison across regions, by patient demographics (age, sex), or over time as a series of cross-sectional estimates. Although many patterns can be identified through indicators stratified in this manner, making sense of the patterns often requires information from other service use indicators as well as information on pathways of care at an individual level.

[1] Corresponding author. E-mail: guido.powell@mail.mcgill.ca

Population health and healthcare information systems rarely support the efficient exploration of hypotheses across multiple regional indicators or temporal patterns within individuals. Understanding how yearly cross-sectional estimates, such as regional distributions of GP visits, relate to similarly stratified estimates of other relevant indicators can be challenging without appropriate approaches to analyzing and visualizing the relationships. For example, regional patterns of GP use may be related to patterns of specialist services, emergency department visits, and hospitalizations in a heterogeneous fashion, both cross-sectionally and longitudinally. Analyses of such health service use at the individual level, may reveal more heterogeneity and complexity by considering the temporal dynamics of various health service use. New strategies for the analysis and visualization of health indicators are clearly needed to support the discovery and interpretation of multidimensional and longitudinal patterns in health indicators.

Our objective is to explore two approaches for analyzing and visualizing indicators of health service use that take into account their multivariate nature, their inherent heterogeneity, and their complex evolution over the temporal progression of an individual disease course. The first approach offers a visualization of multiple indicators of health service use. The approach deals with the complexity of identifying patterns both within and across regions by clustering the data into latent regional profiles. The second approach provides a model of how patients transition through different latent patterns of health service use, based on the same set of indicators, to illustrate variation in patient trajectories over time. Ultimately, these approaches should help decision-makers to identify and interpret patterns within the health system and facilitate data-driven and evidence-based decision making.

2. Methods

This research was conducted as part of the Population Health Record (PopHR) project [3]. PopHR is a semantic web application for measuring and monitoring population health and health system performance. It combines electronic health records on service use, survey data on risk factors of disease, and other types of population health and health system information. The public health insurance provider in Quebec, Canada provided the data on health service use. PopHR's current version uses an open cohort of approximately 1 million people, created by regularly updating a 25% random sample of Montreal's census metropolitan area population.

The analyses were performed on 81,500 patients, over 35 years of age, identified as suffering from chronic obstructive pulmonary disease (COPD), based on diagnostic codes for hospitalization or medical billing (e.g. ICD9 491x, 492x, 496x; ICD10 J41-J44). Information from health services use records was used to develop four specific indicators. We analyzed two types of visits within an outpatient setting, on dates that do not fall within a hospitalization period: visits with a GP and visits with a specialist (coded as respirologist or internist). Visits to an emergency department were counted if occurring outside of a hospitalization period or on the first day of a hospital admission. Finally, we also counted distinct hospital stays, regardless of the responsible diagnosis for hospitalization (all-cause hospitalization). Health service uses were not considered if they occurred prior to the medical events used to identify the COPD cohort.

For the first analysis, each type of indicator was calculated based on the number of COPD patients using each service in the past year as the numerator, over the number of

COPD patients in the corresponding stratum as the denominator. These estimates were aggregated by the regions associated with the 57 local community service centres (CLSC) within the Montreal CMA, standardized for age and sex distributions of the region, and averaged over the period of 2012-2014. The second analysis modeled yearly counts of each service use from 1998-2014, for each of the 81,580 COPD patients.

The first analysis performed a hierarchical cluster analysis on the Euclidean distances of CLSCs four centred indicator values (proportion of COPD patients visiting a GP, a specialist, an ED, or hospitalized), implementing the hclust function in *R*. The intergroup dissimilarity of clusters was used to decide on the number of clusters.

The second analysis fit a hidden Markov model to individual-level aggregates of yearly health service use, using the depmix function from the "depmixS4" package in *R*. The model assumed a Markov process between four discrete latent states of health service use, each defined as five dimensions of service uses (GPs, specialists, ED, hospitalization, or no use). The probability of each service use within a state and probabilities of moving from one state to another were parameters learned by the model.

3. Results

The first cluster analysis using scaled values of each indicator produced visually informative results, identifying multivariate five regional clusters on the four indicators (Figure 1). Without clustering, unscaled mean proportions across the 57

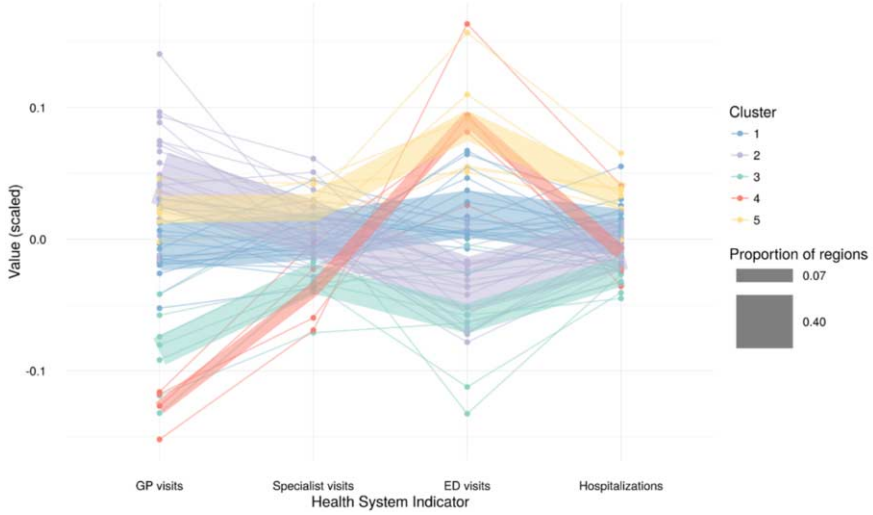


Figure 1. Clusters of 57 health regions in Montreal, Quebec based on the patterns across health system indicators within each region. Color represents the five clusters and the bolder lines, weighted by the size of the cluster, follow the cluster centroids over the indicators.

regions were 0.57 [95% CI: 0.561-0.77] for GP visits, 0.31 [0.311-0.314] for specialist visits, 0.30 [0.288-0.302] for ED visits, and 0.11 [0.112-0.114] for hospitalizations. Clustering revealed two "average" clusters (17 in cluster 1 and 23 in cluster 2), and split 11 regions with low GP use into seven 'low overall' regions (cluster 3) and four 'high ED' regions (cluster 4). Six regions clustered as 'high overall' use (cluster 5).

In the second analysis, the HMM identified four latent health service use states with expected probabilities of service use shown in Table 1. State 1 is characterized by higher specialist visits, state 2 by higher GP visits, state 3 by higher ED visits and hospitalizations, and state 4 by overall low service use.

Patients were most likely to initiate in state 4 (low overall use, 0.64), compared to a 0.18 probability of starting in state 1 (high specialist), 0.08 for state 2 (high GP), and 0.11 for state 3 (high ED and high hospital). On any given year, patients mostly remained in the same state rather than transitioning. Otherwise, patients were most likely to transition to state 4 (low overall). State 3 (high ED and high hospitalization) was the least stable state (i.e. where patients were least likely to remain).

Table 1. Probabilities of observable service use characterizing four latent health service use states in COPD

State	GP visits	Specialist Visits	ED visits	Hospitalizations
1	.008	**.010**	.002	.001
2	**.031**	.002	.004	.001
3	.009	.003	**.017**	**.005**
4	.008	.000	.001	.001

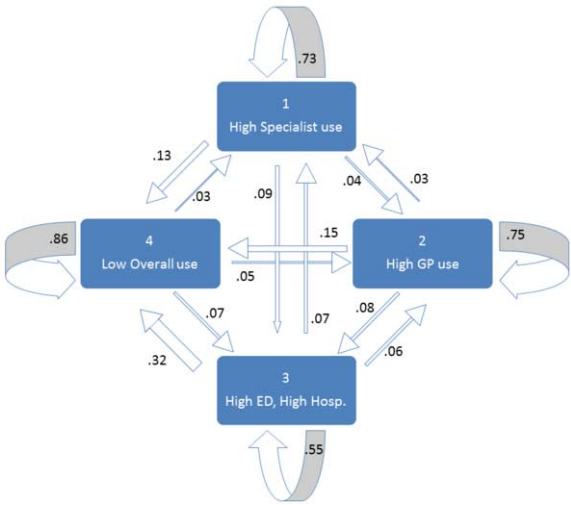


Figure 2. Hidden Markov model of COPD service use. Blue boxes represent hidden health service use states , while arrow direction and width represent transitions between states and their probabilities.

4. Discussion

Our work describes approaches demonstrating the potential for decision-makers to gain insights into health system dynamics through the multidimensional assessments of regional performance and through models of patient trajectories of health care use over time. A population health information system that allows for an interactive implementation of such methods would allow users to better identify regions requiring further examination and help them understand how cross-indicators association are explained by patient trajectories. In contrast to analyses that rank regions or identify outliers on one important indicator, a multivariate clustering approach can give meaning to several regions where estimates estimates of GP use is relatively low. Decision makers can focus interventions where low GP use is associated with high ED

visits, as these estimates are likely more representative of inappropriate care. The second approach we incorporated additional data, summarizing not only the patterns of use of various health service, but also the patient pathways across types of health services. The tendency for COPD patients to predominantly be in a state of low service use is an insight not easily obtained from single, cross-sectional indicators. Nor could a typical indicator have demonstrated that patients are mostly remain in the same state over time. These individual pathways are essential complements to comparisons of regional estimates of different indicators assuming the same population.

Examples of related work in the literature include the use of latent variable models to compare healthcare providers on several correlated dimensions [4]. Research using Markov models and health service data has also modeled progression in various diseases, such as COPD [5]. Our study has demonstrated the value of implementing similar approaches and visualizing results in the context of population health information systems. Though better representing the complexity of the health system, expanding these analysis to include comorbid diagnoses, drugs, or mortality presents challenges to the integration in usable interfaces and in guiding decision making.

The ultimate motivation for monitoring health systems is to improve population health. However, the increasing volume of data on healthcare quality and performance poses certain barriers to effectively make decision to improve health systems. Methods are needed to unify conflicting realities of complex high dimensionality and the need for evidence on which to base decision making. Multivariate and longitudinal analysis and visualization makes better use of the available data and allows for a more comprehensive picture of health systems. These allow useful segmentation of the population and the health system, facilitating the equitable segmentation of resource allocation, and eventually guiding tailored decisions about policy. Better addressing the current and future needs of specific regions and patient groups in such a manner can bring a "precision" perspective to population health [6].

References

[1] W. Hollingworth et al., Using clinical practice variations as a method for commissioners and clinicians to identify and prioritise opportunities for disinvestment in health care: a cross-sectional study, systematic reviews and qualitative study. Southampton (UK): NIHR Journals Library, 2015.

[2] D. Spiegelhalter, C. Sherlaw-Johnson, M. Bardsley, I. Blunt, C. Wood, and O. Grigg, "Statistical methods for healthcare regulation: rating, screening and surveillance," J. R. Stat. Soc. Ser. A Stat. Soc., vol. 175, no. 1, pp. 1–47, Jan. 2012.

[3] A. Shaban-Nejad, M. Lavigne, A. Okhmatovskaia, and D. L. Buckeridge, "PopHR: a knowledge-based platform to support integration, analysis, and visualization of population health data," Ann. N. Y. Acad. Sci., Oct. 2016.

[4] A. Teixeira-Pinto and S.-L. T. Normand, "Statistical Methodology for Classifying Units on the Basis of Multiple Related Measures," Stat. Med., vol. 27, no. 9, pp. 1329–1350, Apr. 2008.

[5] X. Wang, D. Sontag, and F. Wang, "Unsupervised Learning of Disease Progression Models," in Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, NY, USA, 2014, pp. 85–94.

[6] F. S. Collins and H. Varmus, "A New Initiative on Precision Medicine," N. Engl. J. Med., vol. 372, no. 9, pp. 793–795, Feb. 2015.

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-271

Personalized Guideline-Based Treatment Recommendations Using Natural Language Processing Techniques

Matthias BECKER[a,1] and Britta BÖCKMANN[a,b]
[a]*Department of Medical Informatics, University of Applied Sciences and Arts, Dortmund, Germany*
[b]*IMIBE, University Hospital Essen, Germany*

Abstract. Clinical guidelines and clinical pathways are accepted and proven instruments for quality assurance and process optimization. Today, electronic representation of clinical guidelines exists as unstructured text, but is not well-integrated with patient-specific information from electronic health records. Consequently, generic content of the clinical guidelines is accessible, but it is not possible to visualize the position of the patient on the clinical pathway, decision support cannot be provided by personalized guidelines for the next treatment step. The Systematized Nomenclature of Medicine - Clinical Terms (SNOMED CT) provides common reference terminology as well as the semantic link for combining the pathways and the patient-specific information. This paper proposes a model-based approach to support the development of guideline-compliant pathways combined with patient-specific structured and unstructured information using SNOMED CT. To identify SNOMED CT concepts, a software was developed to extract SNOMED CT codes out of structured and unstructured German data to map these with clinical pathways annotated in accordance with the systematized nomenclature.

Keywords. clinical guidelines, information extraction, clinical pathways, knowledge management, natural language processing.

1. Introduction

Clinical guidelines and clinical pathways are used in the healthcare domain to improve the quality of care [1]. The guidelines provide evidence-based medical knowledge around all phases of a disease, while clinical pathways are a process-based standard for treatment of patients within a certain episode. Clinical pathways are structured, multidisciplinary plans of care designed to support the implementation of clinical guidelines and protocols into clinical practise. They aim to improve the continuity and coordination of care across different disciplines and sectors. The transfer of evident knowledge (clinical guidelines) to care processes (clinical pathways) is not a straightforward process because information content and structure differ. It has been demonstrated that in the daily workflow, clinical guidelines provided as real-time decision support systems significantly improve patient care [2, 3] and are effective

[1] Corresponding Author: Matthias Becker, University of Applied Sciences and Arts, Emil-Figge-Str. 42 44227 Dortmund, Germany, Matthias.Becker@fh-dortmund.de

instruments to decrease undesired practice variability [4]. Acceptance of clinical decision support systems is only given if they are integrated into the clinical workflow and presented at the point of care. Model-based approaches to support the development of guideline-compliant pathways, such as PathGuide, enable the formalization of narrative guideline content into care processes [5].

Thus, generic knowledge of clinical guidelines is accessible as defined pathways which can be enacted within different hospital information systems (HIS), but patient-specific information from the HIS is not well-integrated yet. Our goal is to generate personalized support to ascertain the next step within treatment as well as visualize the patient's position on the clinical pathway to identify best subsequent treatment.

To map generic knowledge and patient-specific information, common terminology is needed on both sides. SNOMED CT is an international standard clinical terminology with medical terms and is required to represent the depth and detail of certain procedures and guideline-based care [6]. It provides the reference terminology and the semantic link to combine clinical pathways and patient-specific information. To extract SNOMED CT concepts, a software was developed to identify SNOMED CT codes out of structured and unstructured German data. These were then mapped with pathways annotated in accordance with the systematized nomenclature for personalized guideline-based treatment recommendations.

2. Methods

2.1. SNOMED CT and UMLS

SNOMED CT provides the core general terminology for electronic health records and contains more than 311,000 active concepts. Its comprehensive coverage includes clinical findings, symptoms, diagnoses, procedures, body structures, organisms and other aetiologies, substances, pharmaceuticals, devices and specimens [7]. SNOMED CT is considered the most comprehensive, multilingual, clinical healthcare terminology in the world, but is not available in the German language yet. Therefore, the concept was developed to use English SNOMED-CT Codes by combining them with Unified Medical Language System (UMLS), which is available in German and other non-English languages. This enables mapping of different classifications and ontologies that do not contain the German language [8]. The extraction is achieved through a natural language processing (NLP) pipeline.

2.2. Formalization of Clinical Guidelines to Clinical Pathways

PathGuide is a model-based approach to support the development of pathways compliant with guidelines and the generation of ready-to-use pathway models for hospital information systems and has been developed at the University of Science and Arts in Dortmund, Germany. A meta-model merges the structures of clinical guidelines and clinical pathways into a single generic model. It is encoded through artefacts of Health Level 7 (HL7) in Version 3 [9]. The deployment process to integrate defined guideline-compliant pathways into different target systems is supported by an ontology management approach. This tool enables hospitals to develop guideline-compliant pathways and to integrate them into their HIS without time-consuming manual

transformations. Best practice advice can be provided at the point of care, based on clinical guidelines.

2.3. Information Extraction of Patient-specific Information

In this research, a clinical disorder recognition and encoding system combining a machine learning-based approach for entity recognition with UMLS concept-mapping for the German language is developed. One of the most proven natural language-processing tools is the open-source natural language processing system for extraction of information from electronic medical record clinical free-text Apache cTAKES™ [10, 11]. It already offers a variety of algorithms for text analysis and information extraction. With major modifications, such as the implementation of the German UMLS database and German OpenNLP models, it can normalize to domain ontologies (such as SNOMED-CT) using German UMLS concepts [12, 13]. It processes clinical notes and identifies types of named clinical entities, drugs, diseases/disorders, signs/symptoms, anatomical sites, and procedures.

3. Results

Clinical guidelines provide recommendations for all episodes of a disease including prevention, diagnostics, therapeutics, and rehabilitation. Therefore, the model offers all elements necessary to describe an intersectoral pathway. This information has to be visualized in a transparent manner hiding the complexity of the underlying methods and models wherever possible.

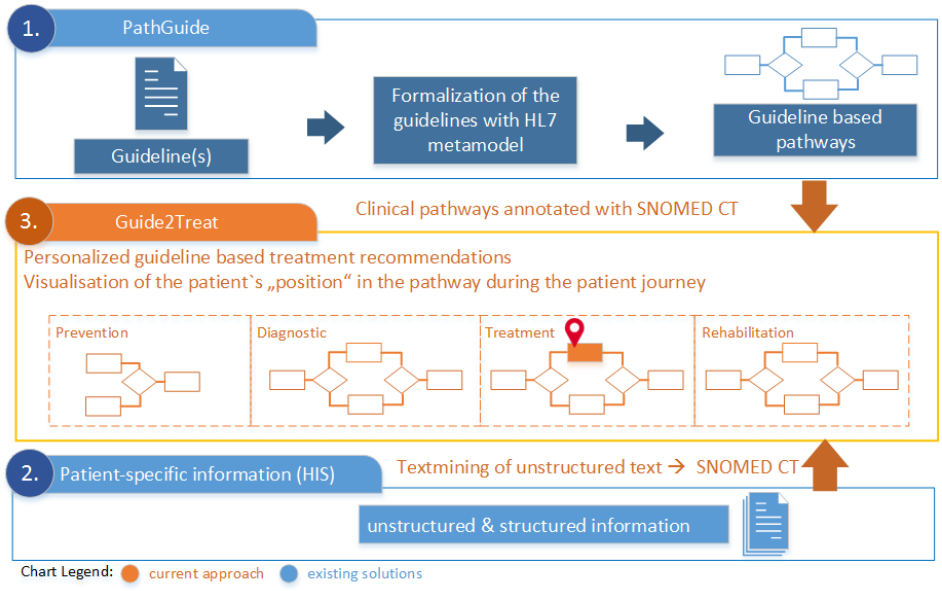


Figure 1. Architecture overview

Figure 1 shows the architecture overview of the personalized guideline-based treatment recommendation system and describes how to manage care pathways by

combining SNOMED CT and patient-specific information held within the HIS. The process of building guideline-based pathways with PathGuide (1.) is defined in three different steps. First, domain experts extract all the pertinent guideline recommendations. The classification is done based on the elements of the HL7-meta-model and indicates the content of a narrative recommendation. The second step includes the formalization of the guideline recommendations. Within the meta-model, all descriptive attributes for a specific element are defined. For this approach, the meta-model is extended using SNOMED CT codes. The last step is the composition of a guideline-compliant pathway by sequencing the activities, defining responsibilities, and adding decision points.

For this extension structured and unstructured information was analysed by the NLP-pipeline (2.). To test and evaluate the pipeline, ShARe/CLEF eHealth 2013 shared Task 1 training set of 199 notes translated in German were used [14]. The evaluation of the pipeline follows the standard metrics of evaluation for the task using F1, i.e., the harmonic mean of Recall and Precision.

Table 1. Statistics of UMLS concepts identified by the NLP-pipeline

Type	#Notes	Recall	Precision	F1
DISCHARGE	61	0.39	0.51	0.44
ECHO	42	0.43	0.74	0.54
RADIOLOGY	42	0.40	0.61	0.48
ECG	54	0.41	0.41	0.41

Table 1 shows the identified UMLS concepts by the NLP-pipeline. 'Precision' is the number of correct positive results divided by the number of all positive results, and 'Recall' is the number of correct positive results divided by the number of positive results that should have been returned. The F1 score can be interpreted as a weighted average of the Precision and Recall, where an F1 score reaches its best value at one and worst at zero. The results of the approach achieved a cautious model with low Recall and higher Precision.

Guide2Treat (3.) connects the components (generic knowledge and patient-specific information) with semantically rich links in a real-time dynamic manner at the point of care and provides the graphical interface for visualization of the patient's position on the pathway. Because the results - especially the recall of the pipeline - are very weak, a fully automatic treatment recommendation is not possible. Therefore, Guide2Treat offers a preselection of patient-specific information. For further patient information, it offers a manual input for the extension of patient information.

Mapping the diagnoses, procedures and findings from the HIS to SNOMED CT enables the SNOMED CT code to work as a parameter, which performs a precise look-up in the system of guidelines. Furthermore, a specific node in the pathway may be connected to a specific information in the HIS, or even to a specific diagnosis, procedure, or finding. By combining the generic knowledge and patient-specific information, it is possible to not only identify the patient's current position, but also the change in positions in the guidelines during the patient's journey. By adding this information, it is possible to create real-time treatment recommendation on the basis of the generic knowledge of the clinical pathways at the point of care.

4. Discussion

There are two main problems with the generation of treatment recommendation: the results of the NLP-pipeline are inaccurate and identification of SNOMED CT codes is

incomplete. The primary challenge is the unavailability of focused international standard SNOMED CT German. Therefore, in comparison to a translated version of SNOMED CT, errors in translation and mapping are more likely. The mapping between UMLS and SNOMED CT depends on the number of entries in the German UMLS database. The identification of SNOMED CT, thus, is only as good as the recognized UMLS concepts and the implemented mapping in cTAKES between UMLS and SNOMED CT. For this reason, to improve the results of the pipeline, the next step is to implement a self-learning UMLS database for the German language and maybe stick with UMLS or, for scale, use of an alternative corpus such as MIMIC-III [15]. As SNOMED CT does not support even such a major language as German, it is questionable whether it is actually worth using it as common reference terminology.

To identify the patient's current position and subsequent changes in positions in the guidelines, more information (e.g. date of operation) is essential. This information cannot be mapped to SNOMED CT and, therefore, a precise look-up in the system of guidelines is not possible, e.g., operation date. Therefore, Guide2Treat has to manage and map more than just SNOMED CT codes to the guideline-compliant pathways. Future work will focus on this mapping challenge and the optimization of the NLP-pipeline.

References

[1] Vanhaecht K: The impact of clinical pathways on the organisation of care processes, *PhD thesis.* University Leuven, Faculty of Medicine; 2007.

[2] Kawamoto K et al. Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. *BMJ.* 2005 Apr 2;330(7494):740-1).

[3] Quaglini S, Ciccarese P. Models for guideline representation. *Neurol Sci*. 2006 Jun;27 Suppl 3:S240-4

[4] Lenz R, Blaser R et al. IT support for clinical pathways--lessons learned. *Stud Health Technol Inform*. 2006;124:645-50

[5] Böckmann B, Heiden K., PathGuide - model-based generation of guideline-compliant pathways for the use in different hospital information systems. *Stud Health Technol Inform.* 2013;192:1089.

[6] Bernstein K., Andersen U., Managing care pathways combining SNOMED CT, archetypes and an electronic guideline system. *Stud Health Technol Inform*. 2008;136:353-8.

[7] T. Benson, *Clinical Terminology, Principles of Health Interoperability HL7 and SNOMED (Health Information Technology Standards)*, Springer, 2012, pp. 201-212

[8] McInnes BT, Pedersen T, Carlis J. Using UMLS Concept Unique Identifiers (CUIs) for word sense disambiguation in the biomedical domain. In: AMIA Annual Symposium Proceedings, Volume 2007. *American Medical Informatics Association:* (2007), 533

[9] Heiden K, Böckmann B. Structured knowledge acquisition for defining guideline-compliant pathways. *Stud Health Technol Inform.* 2013;186:73-7.

[10] Apache, Apache cTAKES, http://ctakes.apache.org/, last access: 15.10.2016.

[11] Savova GK, Masanz JJ, Ogren PV, Zheng J, Sohn S, Kipper-Schuler KC, et al. Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications. *Journal of the American Medical Informatics Association* (2010), 17(5):507–13

[12] Becker M, Böckmann B. Extraction of UMLS® Concepts Using Apache cTAKES™ for German Language. *Stud Health Technol Inform*. 2016;223:71-6.

[13] Divita G, T Zeng Q., Gundlapalli A., Duvall S., Nebeker J., Samore M., Sophia: A Expedient UMLS Concept Extraction Annotator, *Journal of the American Medical Informatics Association* (2014), 467–476

[14] CLEF, ShARe/CLEF eHealth 2013 shared task, https://sites.google.com/site/shareclefehealth/data, last access: 15.10.2016.

[15] Johnson AEW, Pollard TJ, Shen L, Lehman L, Feng M, Ghassemi M, Moody B, Szolovits P, Celi LA, and Mark RG: MIMIC-III, a freely accessible critical care database, *Scientific Data (2016)*. DOI: 10.1038/sdata.2016.35. http://www.nature.com/articles/sdata201635

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-276

HTP-NLP: A New NLP System for High Throughput Phenotyping

Daniel R. SCHLEGEL [a,1], Chris CROWNER [b], Frank LEHOULLIER [b] and Peter L. ELKIN [b]

[a] *Department of Computer Science, SUNY Oswego, Oswego, NY, USA*
[b] *Department of Biomedical Informatics, University at Buffalo, Buffalo, NY, USA*

Abstract. Secondary use of clinical data for research requires a method to quickly process the data so that researchers can quickly extract cohorts. We present two advances in the High Throughput Phenotyping NLP system which support the aim of truly high throughput processing of clinical data, inspired by a characterization of the linguistic properties of such data. Semantic indexing to store and generalize partially-processed results and the use of compositional expressions for ungrammatical text are discussed, along with a set of initial timing results for the system.

Keywords. high throughput phenotyping, clinical NLP, compositional expressions

1. Introduction

Secondary use of clinical data for research requires a method to quickly process large amounts of data for cohort extraction. This is the foundation of phenotyping as the word is used in informatics. More concretely, phenotyping has been defined as "the algorithmic recognition of any cohort within an EHR for a defined purpose, including case-control cohorts for genome-wide association studies, clinical trials, quality metrics, and clinical decision support" [5]. The phenotyping task must be repeated for new versions of semantic resources. A system which performs phenotyping of large volumes of data must be able to do so quickly, and so we often refer to "high throughput" phenotyping.

The general technique for high-throughput phenotyping is to pre-process records to extract information salient to cohort selection. The task is one of information extraction – at least a subset of the contents of records are "understood" and mapped back to a well-defined semantics. Spans of text with identical clinical meaning must be indexed together. Queries may be structured and access this semantic index directly, or unstructured and processed through the same information extraction process, then matched with the contents of the index. These queries may be manually created using Boolean logic on features in the index, or generated automatically by machine learning algorithms, as in [10]. Algorithms for phenotyping are being collected by resources such as PheKB [4].

The term "High Throughput Phenotyping" has been used in this context since at least 2013, when the SHARPn consortium used a template-based method for extracting features related to medications, procedures, symptoms, labs, disorders, and anatomical

[1] Corresponding Author: Daniel R. Schlegel E-mail: daniel.schlegel@oswego.edu.

sites [5]. The system discussed here attempts to be more general, extracting many of the features discussed by the SHARPn consortium, but using much more general templates. Proposed solutions to the phenotyping problem have been around since much before 2013 (e.g., [1]), though the problem itself was not as well defined.

We present two advances in the new High Throughput Phenotyping Natural Language Processing (HTP-NLP) system.[2] Both advances are meant to support high throughput, inspired by a characterization of the linguistic properties of clinical data (Section 2). The HTP-NLP system makes use of a *semantic index* to store partially processed text which may be generalized and re-used (Section 3), and a new implementation of *compositional expressions*[3] for high-speed noun-phrase relation extraction, even from ungrammatical text (Section 4). Initial results are presented in Section 5.

2. Characterizing Unstructured EHR Data

Patient records have properties that are not common to all natural language text, and which contrast with the assumptions of most NLP systems. Records are patient-centric; highly conventionalized; often contain pseudo-structured text which is difficult to analyze; and have text in which the semantic scope is likely to be very local.

Medical records are centered around the patient. In narrative text it is important to track who has a given property, such as in "Susan and Joe met at the hospital. Susan has diabetes and Joe has sleep apnea." where the correct relationship is between "sleep apnea" and "Joe" as opposed to "sleep apnea" and "Susan". In patient records, outside of sections concerning family history, when "sleep apnea" is seen, the assumption is that the condition refers to the patient.

The natural language that occurs in patient records is far less likely to be novel than what is in many other kinds of natural language data. Phrases such as "arrhythmia of the heart" are likely to be frequently repeated, whereas a single phrase in a newspaper may appear extremely rarely. Indeed, patient records use a kind of formalized language. Health professionals establish a convention for expressing things repeatedly, and not always in a grammatical way. A pseudo-sentence such as "EKG: normal." might appear often in fields that are designated for natural-language text, simply because clinicians have established this as the best way to write that a patient's EKG is normal. Another source of conventionalized expressions is the data entry tools used to produce patient health records. A given tool might include a drop-down box with a specific phrase like "cancer of the lung", leading to its repeated appearance.

In many natural language texts the semantic scope of expressions extends over several sentences. The most obvious example of this is pronominal co-reference, where the meaning of a pronoun is taken from the meaning of the previous sentences. In patient medical records cross-sentential semantic relations occur rarely; expressions such as "The patient was diagnosed with skin cancer in 2012. After two years he recovered.", where both sentences are necessary in order to interpret the word "recovered", are rare. The most common type of wide scope encountered in patient medical records is negation, *e.g.,* "The patient does not have skin cancer." The scope of negation is almost always at the level of the sentence rather than having a wider scope.

[2]This software is made available to other CTSA sites, contact elkinp@buffalo.edu for details.

[3]First discussed in preliminary form in [2].

3. Semantic Indexing

The overall design of the HTP-NLP linguistic processing mechanism has been influenced by the linguistic features discussed in Section 2 and the task at hand — quickly building study cohorts. The system makes use of a series of cascading indexes, capitalizing on the highly repetitive and formulaic nature of the natural language data, while preserving the flexibility to adapt to differing study needs.

NLP applications usually processes each document through a pipeline, often without storing intermediate information. These pipelines include steps such as tokenization, sentence breaking, syntactic analysis, and named entity recognition. For a given document, these tasks are performed from left-to-right, top-to-bottom, regardless of how many times the same sentence has been processed. A well designed NLP system which follows this strategy may cache some named entities, for example, to prevent re-processing, but this cache is often limited to the document, and does not persist across an entire corpus.

The HTP-NLP system stores each level of linguistic analysis in a key-value store. Each key represents the incoming level of linguistic analysis and the value represents the outgoing level. One incoming level of linguistic analysis might be the sentence and the outgoing level of linguistic analysis might be noun-phrase chunks. Because of this, each unique sentence and each unique noun-chunk needs to be analyzed only a single time across an entire corpus. Given the formulaic and redundant nature of EHR data, a great deal of redundant processing can be avoided. For example, the phrase "atrial fibrillation" is only coded a single time, even if it appears in the data several hundred thousand times.

The present system uses the following indexes: a discourse level index from individual patient records to sentence/sentence fragments; a syntax level index that from sentences to phrases, words, negation markers, evidentiality markers etc.; a semantic index from phrases or words with polarity indicators to semantically related alternative phrases and synonyms; and an analytic index from linguistic semantic content to ontological specifications such as SNOMED CT, ICD 10, or study-specified ontology. Cohort selection simply requires back-tracking through these indexes from code to document.

This design allows for efficient re-coding in the case of different ontology requirements or updates to existing ontologies. Re-analysis needs only be performed at the level of a single index and only a single time for each indexed key, rather than over each document. Thus, if a study needed to augment an existing index with terms from the Gene Ontology, only the Gene Ontology terms in the index must be re-processed.

4. Resilience to Non-Grammatical Text[4]

Many NLP systems, such as cTAKES [6], rely on some sort of linguistic parse of text for relation extraction. We believe that for at least some cases in the information extraction task, noun-phrase processing can be done using the order of terms as they appear in the text. Relying only on term order allows for processing of ungrammatical text such as "EKG: Normal" and the recognition that it has the same meaning as other expressions such as "normal EKG". Many systems are unable to handle non-grammatical text without specific templates. Our system performs this processing based on a special kind of post-coordination known as compositional expressions (CEs).

[4]Portions of this section adapted from [7]

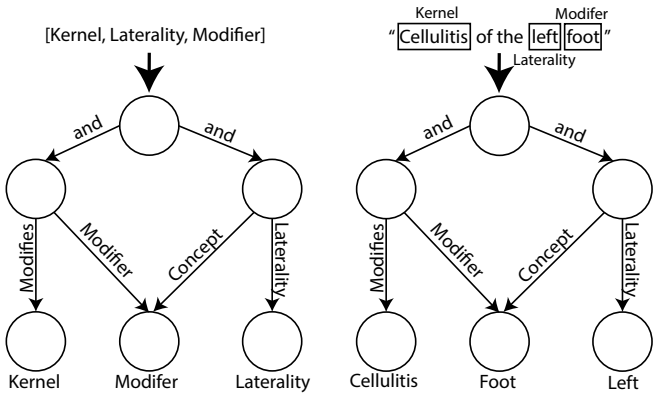


Figure 1. Left: a template graph for text which matches the pattern of a Modifier followed by Laterality followed by Kernel. Right: Instantiation of the template graph by the text "Cellulitis of the left foot".

Post-coordinated concepts consist of multiple individual concepts related to each other using both existing portions of the terminology or ontology graph, as well as new instances of pre-defined relations. Post-coordination is a laborious manual task, in many cases requiring deep understanding of the text and the terminology. For many tasks, such as information extraction, where the goal is to recognize that two spans of text mean the same thing clinically, the deep understanding of post-coordination is often unnecessary.

Compositional expressions extend the idea of using portions of the existing ontology and terminology graph, adding logical and linguistic relations. The advantage of using CEs created from ontologies and terminologies is that multiple surface forms for the same concept are mapped to a single logical form (and hence, a graph structure [8]). For example, the following three forms, all representing hypertension which is uncontrolled, map to a logical form in which the SNOMED CT concept for "hypertension" is the first argument in a binary `hasModifier` relation with the SNOMED CT concept for "uncontrolled": *Uncontrolled hypertension*; *HT, uncontrolled*; *Uncontrolled hypertensive disorder*. In addition, CEs add semantic data which is otherwise missing when text is coded using pre-coordinated terms alone. For example, when using SNOMED CT, 41% of clinical problems require CEs in order to be represented properly [3].

Rules for building CEs are based on the order of terms as they appear in the text. Some rules are based on upper level terms, *e.g.,* a `Procedure` adjacent to a `Body Structure` indicates the site of a procedure. Others use the *kind* of a term – either kernel, modifier, qualifier, or laterality. The kernel is the clinical concept under discussion. Modifiers change the meaning of of the kernel, such as "uncontrolled" in "uncontrolled hypertension." Qualifiers specify status, such as "history." Laterality has to do with sidedness, such as "right." An example of this type of ordering would be a kernel, followed by laterality, followed by a modifier in "cellulitis of the left foot".

Compositional expressions have been implemented to support high throughput. Term orderings are compiled into a discrimination tree, with the leaves mapped to small graph templates. (See left side of Figure 1). A template is instantiated when a sequence of terms matches a path in the discrimination tree, resulting in a graph such as in the right side of Figure 1. These are stored in a graph database, which has been shown to have fast retrieval times [7] for small graphs and subgraphs.

5. Initial Results

To compare processing speed with cTAKES, we processed 537,157 encounter notes for 97,964 patients from the UBMD Allscripts database on a single CPU. This took 48.3 minutes.[5] Of this: 29.3 minutes were spent on linguistic analysis and semantic indexing; 19 minutes were spent coding with SNOMED CT and synonym sets (*e.g.,* [9]), extracting 796 million codes. This is more than an order of magnitude faster than single-threaded cTAKES, which processed 14,021 notes per hour. This improvement is largely attributable to semantic indexing reducing the amount of data which was processed.

6. Conclusion

The HTP-NLP system represents a marked improvement over traditional pipeline-based models for information extraction such as cTAKES. The improvement comes from a system design based on a characterization of the unstructured EHR data which is being processed. The use of semantic indexing results in significant processing-time improvements, and CEs perform high-speed relation extraction, even for ungrammatical text.

Acknowledgements

This work has been supported by the Clinical and Translational Science Award (Number 1UI1TR001412-01), and the VA Big Data-Science Training Enhancement Program.

References

[1] Brown, S.H., et al.: eQuality: electronic quality assessment from narrative clinical reports. In: Mayo Clinic Proceedings. vol. 81, pp. 1472–1481. Elsevier (2006)
[2] Elkin, P.L., Brown, S.H., Chute, C.G.: Guideline for health informatics: Controlled health vocabularies-vocabulary structure and high-level indicators. Stud Health Technol Inform (1), 191–195 (2001)
[3] Elkin, P.L., et al.: Evaluation of the content coverage of SNOMED CT: ability of SNOMED clinical terms to represent clinical problem lists. In: Mayo Clinic Proceedings. vol. 81, pp. 741–748 (2006)
[4] Kirby, J.C., et al.: PheKB: a catalog and workflow for creating electronic phenotype algorithms for transportability. J Am Med Inform Assoc p. ocv202 (2016)
[5] Pathak, J., et al.: Normalization and standardization of electronic health records for high-throughput phenotyping: the SHARPn consortium. J Am Med Inform Assoc 20(e2), e341–e348 (2013)
[6] Savova, G.K., et al.: Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications. J Am Med Inform Assoc 17(5), 507–513 (2010)
[7] Schlegel, D.R., Bona, J.P., Elkin, P.L.: Comparing small graph retrieval performance for ontology concepts in medical texts. In: Wang, F., et al. (eds.) Biomedical Data Management and Graph Online Querying: VLDB 2015 Workshops, Big-O (Q) and DMAH, LNCS, vol. 9579. Springer (2016)
[8] Schlegel, D.R., Shapiro, S.C.: Visually interacting with a knowledge base using frames, logic, and propositional graphs. In: Croitoru, M., et al. (eds.) Graph Structures for Knowledge Representation and Reasoning, LNAI 7205, pp. 188–207. Springer-Verlag, Berlin (2012)
[9] Schlegel, D., Crowner, C., Elkin, P.: Automatically expanding the synonym set of snomed ct using wikipedia. Stud Health Technol Inform 216, 619 – 623 (2015)
[10] Yu, S., et al.: Toward high-throughput phenotyping: unbiased automated feature extraction and selection from knowledge sources. J Am Med Inform Assoc 22(5), 993–1000 (2015)

[5] For a fair comparison with cTAKES, we ensured the same features were found in both. As cTAKES has no CE processor, we excluded it from our test. Processing with CEs roughly doubles coding time.

3. Human, Organisational, and Social Aspects

This page intentionally left blank

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-283

Global *e*Health, Social Business and Citizen Engagement

Siaw-Teng LIAW[a,b,c,1] , Mahfuz ASHRAF[a,b], and Pradeep RAY[a,b]
[a] *UNSW Medicine Australia*
[b] *WHO Collaborating Centre (eHealth), Sydney Australia*
[c] *Ingham Institute of Applied Medical Research, Sydney Australia*

Abstract. The UNSW WHO Collaborating Centre (WHOCC) in eHealth was established in 2013. Its designated activities are: mHealth and evidence-based evaluation, including use case analyses. The UNSW Yunus Social Business Health Hub (YSBHH), established in 2015 to build on the Yunus Centre/Grameen Bank eHealth initiatives, added social business and community participation dimensions to the UNSW global eHealth program. The Grameen Bank is a social business built around microcredit, which are small loans to poor people to enable them to "produce something, sell something, earn something to develop self-reliance and a life of dignity". The vision revolves around global partnerships for development, Millennium Development Goals (MDGs) and Sustainable Development Goals (SDGs). The scope includes mHealth implementation and evaluation in the context of the Internet of Things (IoT), with a growing focus on social business and citizen engagement approaches. This paper summarises a critical case study of the UNSW WHOCC (eHealth) designated activities in collaboration with Bangladesh institutions (International Centre for Diarrhoeal Disease Research, Bangladesh (ICDDRB) and Yunus Centre). Issues and challenges are highlighted..

Keywords. Global eHealth, mHealth, social business, citizen engagement

1. Introduction

The World Health Organisation Collaborating Centre (WHOCC) in eHealth was established in UNSW Medicine in 2013, to implement evidence-based evaluation of eHealth and guide mHealth solutions, including the Internet of Things (IoT). The Yunus Social Business Health Hub (YSBHH) was established in 2015 to build on the Yunus Centre/Grameen Bank eHealth initiatives. The vision revolves around global partnerships for development, United Nations Millennium Development Goals (MDGs) and Sustainable Development Goals (SDGs). The scope includes implementation and evaluation of integrated information systems and data and the Internet of Things (IoT).

The MDGs [1, 2] aim to eradicate extreme poverty and hunger; achieve universal primary education; promote gender equality and empower women; reduce child mortality; improve maternal health; combat HIV/AIDS, malaria, and other diseases; ensure environmental sustainability; and develop a global partnership for development. The SDGs replaced the MDGs in 2016 [3]. Relevant SDGs include:

#8. Decent, full and productive Work and Economic Growth;

[1] Corresponding author, Academic General Practice Unit, School of Public Health & Community Medicine, UNSW Australia, Randwick, NSW 2049 Australia. E-mail: siaw@unsw.edu.au

#9. Resilient and sustainable Industry, Innovation and Infrastructure
#10. Reduced income inequality within and among countries
#11. Sustainable Cities and Communities - inclusive, safe, resilient and sustainable
#12. Responsible and sustainable Consumption and Production patterns

This paper is a critical case study of the WHOCC (eHealth) activities and social business approaches to global eHealth in collaboration with Bangladesh institutions.

The social business paradigm was initiated by Nobel Peace Prize Laureate Professor Yunus as part of the theory and practice of microcredit/microfinance. The Grameen Bank, or "*village bank for the poor*", is a social business built around microcredit or tiny loans to poor people. There is no collateral; the microcredit system is based on trust with no legal documents involved! Grameen Bank now has 9 million borrowers who are "owners" of the bank. Currently, Grameen Bank lends out over one and a half billion US dollars each year. Almost all (97%) borrowers are women who use the loan to start a business: "by producing something, selling something, earning something, she starts to develop self-reliance and a life of dignity" [4].

Social business is defined as a non-dividend business entity established to solve human problems [5]. It can co-exist with money-making business. Successful social businesses in health include selling vegetable seeds at affordable prices to make vegetable growing easy for the citizenry. This business has become the largest seed retailer in Bangladesh and, more importantly, is associated with a marked reduction of night blindness, a common disease among the poor children in rural Bangladesh. Malnutrition is being addressed by a joint venture in 2005 with a global company (Danone) to establish a social business to manufacture an affordable fortified yogurt for poor families. This successful social business continues as part of Danone's corporate social responsibility within a for-profit construct.

The key characteristics of a social business are that it must address a social problem and poverty, operate sustainably, and re-invest profits to expand business and social goals. The social business company adapts business principles to use a market-based solution to alleviate issues caused by poverty, poor health, unhealthy food, smoking, alcohol, gambling, risky behaviour, unemployment, poor literacy, etc. Most of the successful social businesses created in Bangladesh are owned and managed by non-profit organizations. Multi-national companies and non-profits tend to own joint ventures, which may or may not be social businesses.

The Grameen Village Phone Program, started in 1997, provided a good income-earning opportunity to more than 210,000 mostly women Village Phone operators living in rural Bangladesh through facilitating universal access to telecommunication services by the poor in remote, rural areas. The phone was used mainly for financial discussions and social calls with family and relatives living and working in urban areas, resulting in real savings through avoidance of and reduction in trips to the city [6].

In 2006, Grameenphone initiated *HealthLine 789* for its 10 million subscribers, who are charged US$0.38/call for 5 minutes. A range of medical information facilities (e.g. SMS-based laboratory reports), emergency and ambulance services, and real-time medical consultations is provided via mobile phones. A panel of skilled health professionals is available 24/7 through the physician's interface; support is provided by a back office and network manager [7] [8, 9].

High smartphone penetration and strong user and patient demand for mobile phone apps are strong drivers for mHealth initiatives [10]. Health professionals often resist this potential power shift to patients and community. Regulations of the industry do not appear to be a barrier, but, uncertainty exists around the lack of data security and

standards. A general barrier is "*discoverability*", where it difficult to discover the required app from among the estimated 100,000 mHealth apps available online.

Community readiness for eHealth is important globally, as in rural Bangladesh [11, 12]. Community members, leaders and healthcare providers would use mHealth tools and services. However, awareness of existing services is low, especially among the poor and less educated. While face-to-face consultations are preferred, the community is attracted by the timely access to qualified healthcare providers, time savings and lower costs associated with mHealth. Low literacy, lack of English language proficiency, lack of trust and technological incapability were barriers to mHealth use. A sense of ownership, evidence of utility, a positive attitude and intention to use mHealth were drivers of adoption of mHealth services. Implementation strategies must focus on gaining trust, through training and support of users. This requires citizen engagement to inform and empower consumers and ensure transparency and accountability.

The key construct in citizen engagement is public participation. Unlike public communication to inform the public and public consultation, public participation is characterised by 2-way flow of communication in an iterative fashion. It involves the public in collaborative ways and emphasises empowerment. However, barriers exist including poverty and a decreased sense of worth associated with disability and disadvantage such as age, female gender and belonging to minority groups [13].

Global eHealth implementation and evaluation requires social business strategies, targeted at both clinical and population issues, underpinned by citizen engagement if they are to succeed.

2. Progress and Challenges in an Environment of Ongoing Change

2.1. Infrastructure and Building Blocks Activities with International Agencies

- Legal and business analysis of e-Authentication and e-Authorisation [14] as part the EU 7th Framework project called AU2EU {www.au2eu.eu).
- Collaborative work on natural language processing to improve quality of routinely collected data as part of clinical practice [15-18].
- Collaborative multidisciplinary projects to establish the infrastructure and tools to support an Internet of Things approach to the smart built environment.
- Collaborative projects with the Australian Collaborative Research Centre on Spatial Information (CRC-SI) to geocode and spatially enhance the eHealth evaluation methodology to understand integrated health neighbourhoods [19].

2.2. Collaborative mHealth Activities

- with International Centre for Diarrhoeal Disease Research, Bangladesh (ICDDRB) to assess community readiness for mHealth [11, 12, 20].
- with EU partners to assess market adoption, cost, maturity and user acceptance of robotic mHealth services for vulnerable groups [21].
- PhD research into cloud-based mHealth systems for disaster management [22].
- PhD research on mHealth for the primary care of cancer patients [23].

- Indian Aboriginal health agencies on the use of tablets for health checks of independent-living elders, within the Silvercare model where a young retired person supported up to ten elderly people in their neighbourhood [24].

2.3. Evaluation Methodologies and Optimizing the Use of EHR Data for Research

- Data analytics of observational and measurement data from electronic health records from Integrated Health Neighbourhoods (IHNs), which are referral networks across primary and secondary care, supported by an informatics infrastructure and record linkage across clinical and population health information systems, traditional research data sets, social media and personalised appliances [19]. Data quality assessment, management and governance [15, 25-30] are ongoing strategic activities to ensure fitness for purpose for comparative effectiveness research to understand variations in quality of data and care and how eHealth improves self-management, equity and access to health care and social capital across health neighbourhoods.
- EHR data is the core longitudinal data source for quality improvement, evaluation and research; supplemented by specific quantitative and qualitative methods and tools.

3. Conclusions

The WHO CC social business and citizen engagement approach to implementation and evaluation gives meaning to global eHealth infrastructure and tools as sociotechnical mechanisms to achieve access, equity, safety, quality and continuity and comprehensiveness of health care. Individual and community readiness of consumers and providers to adopt and use eHealth tools is central to building a community of citizens that use digitally-enhanced social business approaches to innovate, facilitate and sustain the utility and relevance of global eHealth programs. The highly contextual nature of implementation and evaluation is re-affirmed. The expectations and needs of global eHealth are constantly being met and evolving. The WHOCC (eHealth) research and development program must also evolve with ongoing adaptation and innovation to ensure global eHealth programs and tools contribute meaningfully to meeting the SDGs.

References

[1] United Nations. The Millennium Development Goals Report 2015. New York: 2015.
[2] United Nations. Millenium Development Goal 8: Global Partnership. New York: 2015.
[3] United Nations. Sustainable Development Goals: 17 goals to transform our world 2016.
[4] Dowla A. In credit we trust: Building social capital by Grameen Bank in Bangladesh. The Journal of Socio-Economics. 2006;35(1):102-22.
[5] Yunus M. Creating a world without poverty. Social business and the future of capitalism. New York, USA: Public Affairs; 2007.
[6] Richardson D, et al. Grameen Telecom's Village Phone Programme in Rural Bangladesh: a Multi-Media Case Study. TeleCommons Development Group, Can Int Dev Agency, 2000.
[7] Akter S, Ray P, D'Ambra J. Continuance of mHealth services at the bottom of the pyramid: the roles of service quality and trust. Electron Mark. 2013;23(1):29-47.

[8] World Health Organization. mHealth New horizons for health through mobile technologies Geneva, Switzerland: World Health Organization, 2011.
[9] Al Mamoon I, Khan S. Performance Analysis of a Nationwide Telemedicine Call Center. J Telecommunications. 2011;8(2):10-3.
[10] Zhao J, Freeman B, Li M. Can Mobile Phone Apps Influence People's Health Behavior Change? An Evidence Review. J Med Internet Research. 2016;18(11).
[11] Khatun F, Heywood A, Ray P, Bhuiya A, Liaw S. Determoinants of readiness to adopt mHealth in a rural community of Bangladesh. Int J Med Informatics. 2015.
[12] Khatun F, Heywood AE, Ray PK, Bhuiya A, Liaw S-T. Community readiness for adopting mHealth in rural Bangladesh: A qualitative exploration. Int J Med Inform 2016;93:49-56.
[13] Sheedy A, MacKinnon MP, Pitre S, Watling J. Handbook on Citizen Engagement: Beyond Consultation. Canadian Policy Research Networks, 2008 March. Report No.
[14] Ghorai K, Smits J, Ray P, Kluitman M. European Health Data Privacy and eHealth Rules in Aged Care Coordination. Amsterdam Privacy Conference; 2015; Amsterdam, Netherlands.
[15] Jonnagaddala J, Liaw S, Ray P. Impact of data quality assessment on development of clinical predictive models. Studies in health technology and informatics. 2015(216):1069.
[16] Jonnagaddala J, Liaw S, Ray P, Kumar M, et al. Coronary artery disease risk assessment from unstructured electronic health records using text mining. J Biomed Inform. 2015.
[17] Jonnagaddala J, Liaw S, Ray P, Kumar M, Dai H. HTNSystem: Hypertension information extraction system for unstructured clinical notes. Tech & App Art Intell. 2014:219-27.
[18] Jonnagaddala J, Liaw S, Ray P, et al. Identification and Progression of Heart Disease Risk Factors in Diabetic Patients from Longitudinal EHRs. BioMed Res Int. 2015(10).
[19] Liaw S, de Lusignan S. An 'integrated health neighbourhood' framework to optimise the use of EHR data. J Innovation in Health Informatics. 2016;23(3):547-54.
[20] Khatun F, Heywood et al. Prospects of mHealth to improve the health of the disadvantaged in Bangladesh. In Adibi, S (Ed) *mHealth multidisciplinary verticals*. Taylor & Francis 2015.
[21] Ariani A, Kapadia V, Talaei-Khoei A, Li J, Ray P. Challenges in Adoption of Assistive Robots among Seniors. Int Technology Management Review (in press). 2016.
[22] Guempana Y, Rabhi F, Lewis J, Ray P, Zhu L, editors. Mobile Cloud Computing for Disaster Emergency Operations. IEEE ISTAS2015; 2015 Nov; Dublin, Ireland.
[23] Lewis J, Ray P, Liaw S-T. Recent worldwide developments in eHealth and mHealth to more effectively manage cancer and other chronic diseases. Int Yrbk Med Inform. 2016:11-26.
[24] Blake J, Ray P, editors. Facilitating Digital Communication in Seniors. IEEE International Symposium on Technology and Society (ISTAS2016); 2016 Oct 21-22; Trivandrum, India.
[25] de Lusignan S, Liyanage H, Di Iorio C, et al. Using routinely collected health data for surveillance, quality improvement and research. J Innov Health Inform. 2015;22(4):426–32.
[26] de Lusignan S, Liaw ST, et al. Key concepts to assess the readiness of data for International research. Int Yrbk Med Inform. 2011;2011:112-21.
[27] Kahn M, Callahan T, al e. A Harmonized Data Quality Assessment Terminology and Framework for the Secondary Use of Electronic Health Record Data. eGEMS 2016.
[28] Liaw S, Rahimi A, Ray P, et al. Towards an ontology for data quality in integrated chronic disease: a realist review of the literature. Int J Med Informatics. 2013;82(1):10–24.
[29] Rahimi A, Liaw S, et al. Ontological specification of quality of chronic disease data in EHRs to support decision analytics: a realist review. Decision Analytics. 2014;1(1):5.
[30] Taggart J, Liaw S-T, Yu H. Structured data quality reports to improve EHR data quality. Int J Med Informatics. 2015;84:1094-8.

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-288

Project PEACH at UCLH: Student Projects in Healthcare Computing

Navin RAMACHANDRAN[a], Dean MOHAMEDALLY[b] and Paul TAYLOR[b1]

[a] *University College London Hospital, London*
[b] *University College London, London*

Abstract. A collaboration between clinicians at UCLH and the Dept of Computer Science at UCL is giving students of computer science the opportunity to undertake real healthcare computing projects as part of their education. This is enabling the creation of a significant research computing platform within the Trust, based on open source components and hosted in the cloud, while providing a large group of students with experience of the specific challenges of health IT.

Keywords. Capacity building, Informatics education. Open source

1. Introduction

A key challenge for the health informatics community is to train a workforce capable of delivering the potential of new technology. While attention has focused on initiatives to give clinicians training in informatics, we believe that we must also train computer scientists to have a strong understanding of the special requirements of healthcare: the additional responsibilities for patient safety and confidentiality that healthcare IT involves, and the complexity of the environment in which systems have to work.

To put this in a UK context, the NHS workforce now numbers around 1.3 million staff. A 2008 estimate (there isn't a more recent figure) suggested that informatics staff represent 3% of total and that vacancy rates were as high as 16%. [1] Although written evidence is hard to find, other employers tell us that they too struggle to recruit technical staff with specialist knowledge of healthcare.

Five years ago the Department of Computer Science at UCL began to move away from the use of invented or artificial problems in teaching. This initiative led to the development of the Industry Exchange Network (http://ixnet.org.uk/) to train students on real-world problems through term-based projects with real clients. This is now the largest network of client engagement projects for term-based CS students in the country. In 2016 the authors of this paper began what we hope will be a long-running initiative to host a large number of these projects within a hospital. The students are supervised by staff from the university's computer science department, with input from a health informatics specialist. The projects are proposed by a clinical team who act as

[1] Corresponding author, Paul Taylor, Institute of Health Informatics, UCL, 222 Euston Rd, London, NW1 2DA, United Kingdom; E-mail: p.taylor@ucl.ac.uk

clients for the duration of the projects, have regular meetings with the students, and provide guidance and advice throughout.

The projects, collectively known as PEACH, are all based on the use of open source components and, it is hoped, will provide a proof-of-concept for open source software development in the NHS as well as a realistic exposure to the special demands of health informatics.

2. Methods and Participants

Twenty UCL MSc students, five from the specialist Software Systems Engineering MSc and fifteen from the generalist MSc in computer science (designed for students who did not take computer science as a first degree) volunteered to work on a suite of summer projects based at UCLH.

The projects were proposed by the clinical clients: Dr Navin Ramachandran (the first author of this paper and a consultant radiologist) and Dr. Wai Keong Wong (consultant haematologist at UCLH). Projects were identified and students assigned roles within each project. Students worked in five teams and each project was assigned an architect from the students in the SSE MSc.

All students had a supervisor from UCL Computer Science who met with them at the outset (in June 2016), at the mid-project review, at the final assessment (in Sept 2016) and as required in between. The clients met with the students weekly, or more often if required. All students completed online training in information governance and data security for healthcare researchers.

Additional support was supplied by Microsoft UK who attended early meetings, provided a mentor for students and donated Azure credits to support the applications.

Test data for the projects was supplied by UCLH and consisted of 5 years' worth of anonymized data from the hospital RIS system and a set of test data from the Intensive Care Unit provided through the Health Informatics Collaborative.

2.1. Core Platform

A core platform was engineered to support the deployment of multiple healthcare applications. The basis of the platform was an integration and analytics engine composed of the following elements. Apache NiFi, a dataflow tool, was used to import the data from the files supplied by the Trust. Apache Kafka was used both as a messaging hub and a source of truth (in a Kappa architecture).[2] This works on a publish -subscribe model, whereby sources publish data to Kafka and applications subscribe to the data feeds. Druid was used for online analytical processing and Apache Spark for machine learning analytics. The whole platform was deployed on DCOS (data centre OS), hosted on Microsoft Azure, which also provided the required role based security.

For novel healthcare data the client selected EtherCIS as the main datastore. [3] EtherCIS is an open source system, compliant with the openEHR information modeling specification.[4] This system provides an interoperable backend service to applications, allowing them to use RESTful APIs to persist and retrieve healthcare data.

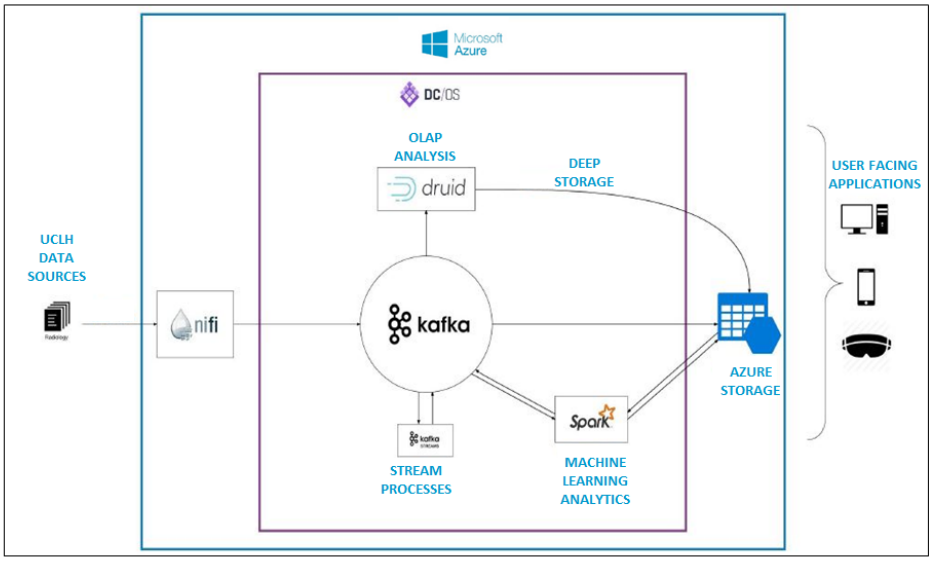


Figure 1. Architecture diagram of the PEACH data platform

2.2. Applications

- The "Multidisciplinary Team (MDT)" application is an innovative application to manage information on patients reviewed during cancer MDT meetings.
- The "Dashboards" application is a customizable reporting tool for multiple users, supporting interactive analysis on event data to provide real time visualization of key statistics, with radiology data as the first use case. Users have access to a variety of services based on roles.
- The "Chatbot" team developed an interactive conversational chatbot using natural language processing and machine learning to perform an electronic Holistic Needs Assessment (eHNA) for cancer patients, in order replace a very time-heavy process.
- The "Reality" application allows for surgeons to interact with 3D models of patients pre-operatively using virtual reality, in order to prepare for surgery. The first iteration of the project considered preoperative planning for renal cancer surgery.
- The "Clinical Trials" application provides a searchable online database of clinical trials in the region, for administrators, clinicians and patients.
- The "Visual Report" application is a novel approach to healthcare data recording, using tablet / pen input to record diagrams of prostate cancers on a template, with analysis of the diagram to provide automated risk scoring of prostate regions.

3. Results

The PEACH project consists of multiple projects, which were assessed separately. Performance of the core platform was measured by its processing time. In Table 1 we provide a sample measurement in the pre-processing layer based on 10,000 records of one of the csv data files (DATAEXAM file). Complexity of each stage of testing was measured by counting the number of processors involved at that stage. The metrics demonstrated that a CSV file could be pre-processed to singular JSON objects in less than 4 milliseconds per record in Stage 2. In Stage 3, the time required to finish a 10,000 record transfer to the message hub cluster was 2 hours and 47 minutes, or less than 1.001 second per record. These figures demonstrate that using a single process to deal with both batch data (large volume) and streaming data (low volume but realtime) in a Kappa architecture was a viable approach.

Table 1. Sample result of measurement in pre-processing layer

Stage	Input files	Input MB	Output files	Output MB	Time	Complexity
1	1 (CSV)	494.31	299 (CSV)	494.31	00:00:21.227	13
2	1 (CSV)	1.65	10,000 (JSON)	7.17	00:00:37.419	9
3	10,000 (JSON)	7.17	10,000 (key,value)	7.17	02:46:50.722	5
2-3	1 (CSV)	1.65	10,000 (key,value	7.17	00:00:50.467	12

The MDT team created a 'Dockerised' version of EtherCIS, a significant achievement. [5] However there were problems with the reliability of solutions for role-based access. The team was unable to demonstrate a working front-end.

The Clinical Trials application team was assisted by experts in clinical terminologies and was able to solve many of the backend challenges in representing trial inclusion and exclusion criteria. One problematic aspect of this project was that the team chose to implement the app using a different and suboptimal framework (sails.js). Future projects will use a single consistent architecture.

The Visual Report team created a working intuitive product which it is hoped could be published on the Windows App store with further development,

The radiology Dashboard team created a real-time batch pipeline using the core platform. Unfortunately the dashboard interface was only partially completed. A major problematic requirement was for a UK data-centre, in order to meet information governance rules - these were all in Technical Preview phase with limited functionality. Microsoft have since provided further assistance to resolve this issue as well as 20,000 USD Azure credits for the next phase of the work.

The Chatbot was less successful due to suboptimal integration between components developed by different team members. Future projects will be stored on a common GitLab repository, enforcing continuous integration, to mitigate this risk.

The Reality project worked well, allowing 3D anatomical reconstructions to be viewed on the HTC Vive, but did not have a fully-developed pipeline for producing the required 3D models. The virtual reality visualization also proved slightly disorientating for users. The next phase will use Hololens Augmented Reality (which should be less disorientating), and also develop a model construction pipeline and design language.

One student wrote of the experience of leading a PEACH project: “It was very useful especially for us because we haven't had any previous experience on patient data before PEACH. It was really challenging to keep it secure especially when the UCL-UCLH sharing agreement process demanded a long time/effort before the data could actually be securely transferred to our server. Our everyday communication with the doctors and healthcare specialists gave us a better understanding of how a real life problem looks like and the kind of influence our project (with its weak and strong points) could have in people's lives. For me, the aspects of security, efficiency and consistency of data across a real time infrastructure were really fascinating. It's like a dream for me to help people through technology and try to solve a real need.”

4. Conclusions

The following lessons were learned and are being applied in the next phase:

- The number of core technologies has been reduced to ensure interoperation and reusability. A team will build a reusable component library.
- Many students had not been sufficiently encouraged to think about teamwork. Some supervisors instructed students to concentrate on their own components and not on the reusability of the application. The guidance has been changed.
- There were a number of tensions between the supervisor’s and the client's requirements. For example the client might ask for a simpler interface where the supervisor may want the student to use demonstrate more complex skills. We will improve communication between supervisors and clients offset this.
- We have improved the guidance at the outset of the project so that students have a clearer statement of requirements and principles of continuous integration and continuous deployment are enforced.
- The cloud services are available from the start.
- We will generate a realistic synthetic dataset to use in testing.

Acknowledgements

Dr Wai Keong, Dr Jay Kola and Dr Ian McNicoll provided input and guidance to the student projects. Efthymia Kazakou and the other PEACH students contributed to the success of the work reported here.

References

[1] Tribal Consulting, “NHS Informatics Workforce Survey 2007/08,” 2008.
[2] J. Kreps, N. Narkhede, and J. Rao, “Kafka: a distributed messaging system for log processing,” in *Proceedings of the NetDB*, 2011.
[3] “EtherCIS by ethercis.” [Online]. Available: http://ethercis.github.io/. [Accessed: 04-Nov-2016].
[4] A. Moreno-Conde *et al.*, “Clinical information modeling processes for semantic interoperability of electronic health records: systematic review and inductive analysis,” *J. Am. Med. Inform. Assoc.*, vol. 22, no. 4, pp. 925–934, Jul. 2015.
[5] D. Merkel, “Docker: Lightweight Linux Containers for Consistent Development and Deployment,” *Linux J*, vol. 2014, no. 239, Mar. 2014.

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-293

Monitoring of Students' Interaction in Online Learning Settings by Structural Network Analysis and Indicators

Elske AMMENWERTH[a,1] and Werner O. HACKL[a]
[a] *Institute of Medical Informatics, UMIT – University for Health Sciences, Medical Informatics and Technology, Hall in Tirol, Austria*

Abstract. Learning as a constructive process works best in interaction with other learners. Support of social interaction processes is a particular challenge within online learning settings due to the spatial and temporal distribution of participants. It should thus be carefully monitored. We present structural network analysis and related indicators to analyse and visualize interaction patterns of participants in online learning settings. We validate this approach in two online courses and show how the visualization helps to monitor interaction and to identify activity profiles of learners. Structural network analysis is a feasible approach for an analysis of the intensity and direction of interaction in online learning settings.

Keywords. Learning; Education, distance; Cooperative behavior

1. Introduction

Learning can be understood as constructive and social process that works best in interaction with other persons [1]. Through interaction and collaboration, students "gradually construct systems of shared meanings" [2]. Studies show advantages of collaboration in learning activities, such as more engaged learning, increased motivation and attention, more active processing of information, improvement of meta-cognitive and social skills and overall better knowledge acquisition and retention [3].

Interaction between students is also considered a key element for successful learning in online settings [4]. However, the teacher needs to address specific challenges of online settings such as reduced transmission of socio-emotional information, more complicated coordination of asynchronous activities and the challenge of lurking, i.e. the more passive participation in online activities [5]. Collaborative online teaching thus needs thoughtful instructional design to facilitate the interactions of the students and a close monitoring of the quantity and quality of interactions.

In this contribution, we present an approach and related indicators to analyse and visualize interaction patterns of participants in online learning settings. We validate this approach in two online courses and show how the visualization helps to monitor interaction and to identify activity profiles of learners. We conclude with recommendations for online teaching.

[1] Corresponding author, Elske Ammenwerth, Eduard Wallnöfer Zentrum 1, 6060 Hall in Tirol, elske.ammenwerth@umit.at.

2. Methods for Analysis and Visualization of Communication Networks

Interaction patterns in online courses can be analyzed in various ways. Analysis may focus on quantitative numbers such as numbers of contributions, continuity of participation, or number of answers in relation to all contribution [2]. Another line of analysis focusses on the content of messages and tries to characterize these e.g. into questions, answers, agreement or disagreement [4]. Graphical visualization of communication patterns between participants in network diagrams can help to identify interaction patterns [2]. Finally, social network analysis methods can be used to statistically analyse e.g. intensity, cohesion or reciprocity of participants [6].

To monitor effects of collaborative online teaching, we were especially interested in analyzing the interaction network of the participants and its changes over time. We used indicators from structural network analysis and a graphical visualization of network activity.

We took two fully online modules as case study. Both modules were run in 2016 with 16 resp. 15 participants from various professional fields within healthcare. The first module went for four weeks, the second for six weeks. The first module focused on project management, the second on clinical data analytics. The participants in the first course were partly familiar with each other, while participants in the second course mostly did not know each other before. Estimated student workload for both courses was 10 – 15 hours per week. Participants received a certificate upon completion.

As instructional theory, we used the expository 3-2-1-design framework by Michael Kerres as a basis and combined it with the concept of E-tivities by Gilly Salmon [5]. Each course consisted of meta-information (on learning objectives, estimated workload, instructional approach etc.) and a set of learning activities. Each learning activity comprised a structured description of learning objectives, tasks to be done and expected reaction to the solutions of other participants. These learning activities are not meant to test competencies, but to allow the students – alone and in interaction with the others – to accomplish the intended learning process. At the end of each week, participants were asked to write a reflection on their learning progress. For each learning activity, the needed materials (presentation, paper, book chapters, or web sites) were provided by the instructor. Moodle was used as electronic learning platform.

Log data from the learning platform Moodle was exported in anonymized form and analysed using the Talend Open Studio software platform (www.talend.com) and Tableau 10.0 (www.tableau .com). Interaction patterns of participants were analyzed and visualized for each week of the module using the graph visualization and manipulation platform Gephi (www.gephi.org).

Interaction indicators were adapted from the structural analysis by Coll et al [2]. From the three basic types of interaction [7], i.e. learner-content, learner-instructor, and learner-learner interaction, we focus on learner-learner interaction. A written survey and interviews with all participants were conducted at the end of each course to explore satisfaction with the course and learning outcome. In addition, a workload analysis was conducted.

3. Results

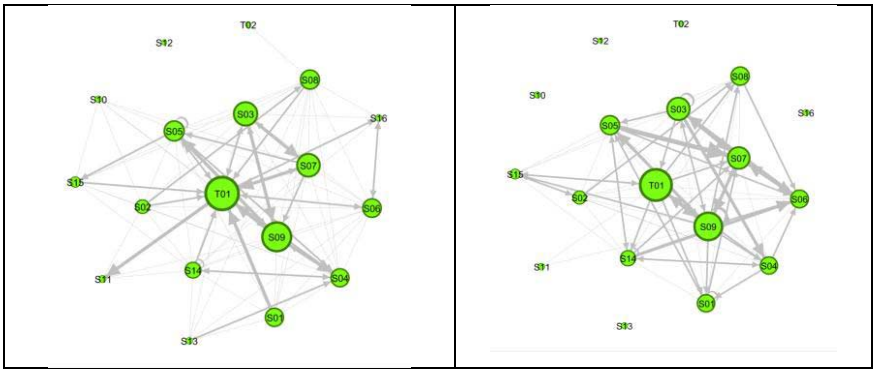
Course 1 comprised 29 learning activities and 29 forums with 287 threads and 1,605 posts altogether. 998 forum posts were written by the successful participants, 178 by

the course instructor. Course 2 comprised 25 learning activities and 25 forums with 153 threads and 1,232 posts altogether. In this course, 702 forum posts were written by the successful participants, 187 by the course instructor. In course 1, 9 of the 14 participants successfully completed the course. In course 2, 8 of 16 participants successfully completed the course. Table 1 shows indicators of these successful participants. The number show that successful participants reach the threshold for two of the each indicators defined by [2]: Access index > 0.5; contribution index > 0.6; answer-contribution index for the two courses was below, but close to the 0.9 threshold.

Table 1. Interaction network indicators for the successful participants of two online courses. "post" = active contribution to one discussion forum.

	Course 1 (n=9) Min | Max | **Mean** | StandardDev	Course 2 (n=8) Min | Max | **Mean** | StandardDev
Access index (% of days online)	0.56 | 0.94 | **0.80** | 0.13	0.64 | 0.92 | **0.81** | 0.09
Activity index (% of days with at least one post)	0.49 | 0.89 | **0.67** | 0.15	0.44 | 0.86 | **0.63** | 0.15
Reading index (% of read posts in relation to all posts)	n.a. due to chosen settings in the electronic learning platform (automatic email-notification)	
Completion index (activities with at least one post)	0.93 | 1.0 | **0.97** | 0.03	0.76 | 0.95 | **0.86** | 0.07
Contribution index (relation of written posts to number of minimum requested posts)	1.241 | 3.052 | **1.91** | 0.602 (58 requested posts)	n.a. (no requested minimum posts)
Thread-starting index (% of threads that were started by a particular user in relation to all posts of successful students)	0.023 | 0.029 | **0.026** | 0.002	0.011 | 0.161 | **0.036** | 0.047
Answer-contribution index (% of post that are answers to other posts in relation to all post of a student)	0.67 | 0.85| **0.75** | 0.07	0.76 | 0.88 | **0.83** | 0.04

Figure 1 shows the interaction network of course 1. In week 1, the instructor is in the middle of interaction. Starting with week 2, the role of the instructor is less important, and the interaction between participants increases. Figure 2 shows that certain interaction profiles – receiving participant, sending participant, and balanced participant – can be identified based on their individual activities.



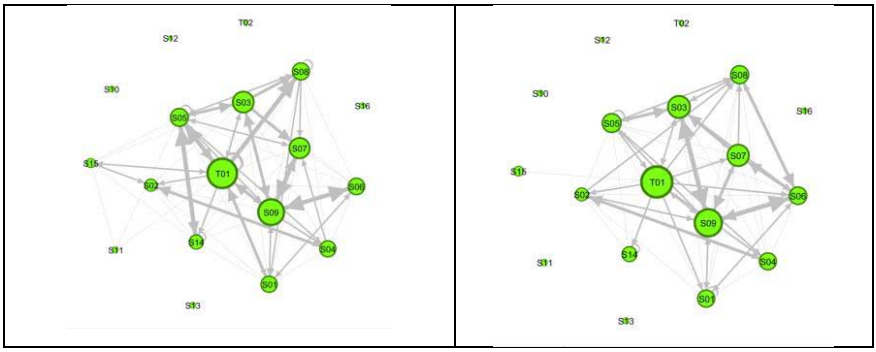


Figure 1. Interaction network in week 1 – 4 in the first course. Circles indicate participants (indicated by Sxx) or the instructor (T01). Size of circles indicates the number of sent messages (larger circle correspond to more posts over the course period). Arrows indicate direction and intensity of communication.

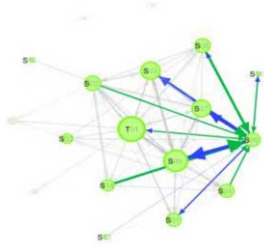
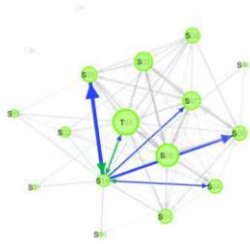
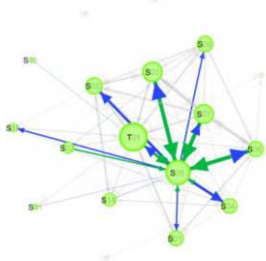
Type of participant	*Example of interaction network*
Receiving participant: Amount of received messages is larger than number of send messages.	
Sending participant Amount of sent messages is larger than number of received messages.	
Balanced participant Amount of received messages and of send messages is nearly equal.	

Figure 2. Various types of participants, with examples of network interaction diagram from course 1.

4. Discussion and Conclusion

Interaction with peers is a prerequisite for learning: "No interaction, no education." [8]. We showed how structural network analysis can help to describe indicators and activity profiles of participants. This information can help to track the activity of participants during the course and to identify those who need teacher support [2]. It also helps to evaluate post-hoc whether the chosen instructional design has led to a collaborative atmosphere within the online course as intended to facilitate learning and to refine the instructional design if needed [9].

Data relating to the learning process of the students in online courses is nowadays easily available through the learning management systems (LMS), but exploitation of this data is still rare [9]. In our case study, the data was extracted manually. For future routine use, automatic procedures need to be developed to allow data extraction and indicator generation.

To assess whether interaction contributes to learning, we will correlate learning outcome with level of interaction. Gunawardena has described how online learners can arrive at a higher level of critical thinking through different stages of interaction with peers: (a) sharing/comparing of information, (b) discovery of dissonance, (c) co-construction of knowledge (d) testing and modification of proposed synthesis, and (e) agreement of newly constructed meaning [10]. The presented indicators and interaction networks do not allow describing in which stage the observed interaction took place. For this, a content analysis of the contributions and posts is needed in addition to the structural analysis. Such an analysis would further contribute to the understanding of online interaction with a focus on a learner-centered instructional design.

References

[1] Vygotzky L, *The Development of Higher Psychological Processes*, Cambridge, Massachusetts: Harvard University Press, 1978.

[2] Coll C, Engel A, Bustos A, Distributed Teaching Presence and Participants' Activity Profiles: a theoretical approach to the structural analysis of Asynchronous Learning Networks, *European Journal of Education*, **44**(4) (2009),512-38.

[3] Rey G, *E-Learning*, Bern: Huber, 2009.

[4] Chou C. A Comparative Content Analysis of Student Interaction in Synchronous and Asynchronous Learning Networks. Proceedings of the 35th Hawaii International Conference on System Sciences. 10.1109/HICSS.2002.9940932002.

[5] Salmon G, *Etivities – The key to active online learning*, New York: Routledge, 2013.

[6] Stepanyan K, Mather R, Dalrymple R, Culture, role and group work: A social network analysis perspective on an online collaborative course, *British Journal of Education Technology*, **45**(4) (2014),676-93.

[7] Moore M, Editorial: Three types of interaction, *The American Journal of Distance Education*, **3**(2) (1989),1-6.

[8] Gunawardena C, Lowe C, Anderson T. Transcript analysis of computer-mediated conferences as a tool for testing constructivist and social-constructivist learning theories. Proceedings: Distance learning '98: The 14th Annual Conference on Distance Teaching and Learning. Madison, WI: University of Wisconsin; 1998. p. 139-45.

[9] Li Y, Bao H, Xu C. Learning Analytics: Serving the Learning Process Design and Optimization. In: Lai F, Lehmann J, editors. Learning and Knowledge Analytics in Open Education. Switzerland: Springer; 2017. p. 31-40.

[10] Gunawardena C, Lowe CA, Anderson T, Analysis of a global online debate and the development of an interaction analysis model for examining social construction of knowledge in computer conferencing., *Journal of Educational Computing Research*, **17**(4) (1997),397-431.

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-298

A Lens for Evaluating Genetic Information Governance Models: Balancing Equity, Efficiency and Sustainability

Espen SKORVE[a,1], Polyxeni VASSILAKOPOULOU[b,] Margunn AANESTAD[c], Thomas GRÜNFELD[d]
[a]*Aalborg University, Denmark*
[b]*University of Agder, Norway*
[c]*University of Oslo, Norway*
[d]*Oslo University Hospital, Norway*

Abstract. This paper draws from the literature on collective action and the governance of the commons to address the governance of genetic data on variants of specific genes. Specifically, the data arrangements under study relate to the *BRCA* genes (*BRCA1* and *BRCA2*) which are linked to breast and ovarian cancer. These data are stored in global genetic data repositories and accessed by researchers and clinicians, from both public and private institutions. The current *BRCA* data arrangements are fragmented and politicized as there are multiple tensions around data ownership and sharing. Three key principles are proposed for forming and evaluating data governance arrangements in the field. These principles are: equity, efficiency and sustainability.

Keywords. Genetic information, governance, the commons.

1. Introduction

The significance of genetic testing for clinical purposes is increasing, and the rapid advances in sequencing technologies are currently amplifying this development. One outcome of this is the exponential growth in genetic data generation. New data are continuously produced as output of genetic analyses performed all over the world. The new data are subsequently used as an essential input for further analyses for research and clinical purposes alike. In this paper, we focus on data related to the (potential) pathogenicity of variants of specific genes. These data relate to single genes, they are completely anonymous, and do not raise any privacy issues. They are valuable for taking clinical decisions related to diagnostics, treatments and prevention [1][2][3]. Their importance for decision taking is making it urgent to address data governance in the domain addressing the tensions, contestations and controversies around data ownership and control. These tensions are shaping the currently fragmented and politicized landscape of gene-specific data repositories. The current situation is making it difficult to reap the benefits of the increased speed and reduced cost associated with

[1] Corresponding Author: Espen Skorve, Department of Computer Science, Aalborg University, Selma Lagerlöfs Vej 300, 9220 Aalborg East, Denmark. E-mail: espesko@cs.aau.dk

new sequencing technologies [4][5]. In order to address this issue, we need a perspective that can be conducive to reaching resolutions within such a complex domain. In this paper, we draw from theoretical resources on "the commons" that can be used as a basis for resolving governance issues and for facilitating a sustainable development within the field.

We turn to the theory of the commons [6][7] aiming to draw concepts for addressing "*the age-old problem of how to induce collaborative problem solving and other forms of collective action among self-interested individuals, groups, or organizations, assuming, of course, that they share at least some common goals*" [8]. In this body of literature, a basic assumption is that different types of resources are subject to different governance considerations, and a common taxonomy for discussing this is drawn along the axis of subtractability and exclusion [9] [10]. Based on a claim that knowledge about the human genome belongs to mankind as a whole, current attempts at privatizing data that can contribute to this knowledge have been characterized as a second enclosure movement [11][2]. We argue that biomedical knowledge in general, and genetic knowledge in particular, is of such a character that it constitutes what we call a public good. Hence, we adopt three fundamental concepts for evaluating any regime aiming to govern data that can contribute this knowledge resource: equity, efficiency and sustainability [10].

2. Case Background and Method

A typical example of the genetic data governance challenges can be found in the specialized field of breast cancer genetics, and the way this has evolved over the two decades that passed since the identification of the *BRCA* genes. The *BRCA* genes (*BRCA1* and *BRCA2*) are related to susceptibility to breast and ovarian cancer. Multiple data repositories containing information regarding the pathogenicity associated with specific variants of these two genes have emerged during that past decades. A main distinction between these repositories is how contribution and access rights are organized and governed, and their differences are most often related to the degree of openness. The Breast Cancer Information Core (BIC) was the first open access repository of *BRCA* data (it was established in 1995). This is a shared repository where information generated from research or clinical practice, within private and public laboratories all over the world can be deposited and retrieved. In the U.S.A, one private laboratory (Myriad Laboratories) was able to establish a dominant position in the local market for *BRCA* testing after having been granted several patents based on its central role in the discovery of the *BRCA* genes. This laboratory was the primary contributor to the BIC database for several years, and the '*BRCA* community' was fairly dependent on Myriad's sharing of knowledge and information. When Myriad decided to discontinue their contributions to BIC in 2004, this was therefore a major event that caused several counter actions aimed at amending the problems arising from significant information being disclosed from the community. Through initiatives like the 'Sharing Clinical Reports' and 'Free the Data' projects, new open access repositories were established, and doctors and patients were encouraged to register the detailed results from their lab tests at Myriad's [12][13]. Furthermore, several other initiatives have emerged on both

[2] The privatization of common land in Europe during the 15th century constitutes the first enclosure movement [11].

sides of the Atlantic. The Leiden University Medical Center (Netherlands) launched 'LOVD' as early as 2005, and 'ClinVar' was released by the US National Center for Biotechnology Information (NCBI) in 2012. In 2015, the 'BRCA Mutation Database' (by the University of Utah's Department of Pathology and the ARUP Laboratories. USA), the 'BRCA Share' repository (INSERM and Quest Diagnostics, France) and the repository 'BRCA Exchange' (by the Global Alliance for Genomics and Health) were launched. The latter is particularly interesting, as it constitutes a global, networked initiative (as indicated by the name), and thus most strongly manifests the general tendency of initiatives the span the boundaries of single, local or even national actors. However, this increasing multiplicity of repositories and initiatives in the domain results in duplication of efforts and increased difficulty in data retrieval.

The research reported in this paper is designed as a case study [14] with focus on the shaping of governance arrangements for information related to the (potential) pathogenicity of *BRCA* variants. We collected empirical material over a two years´ period as part of the activities of a research and development project that aimed to develop a secure IT platform to facilitate distributed collaboration and access to genetic information. We performed 12 interviews with experts in the domain and we reviewed more than 100 documents (journal papers, specialised press reports, commercial announcements) to identify key events that mark the evolution of governance arrangements in BRCA domain. We adopted an interpretive approach for the analysis of the data [15] going through transcripts, notes and documents in order to identify relevant themes. We structured our analysis around the key theoretical concepts of equity, efficiency and sustainability. Initial findings were refined and verified based on continuous communications with practitioners in the field of genetic analysis.

3. Addressing Information Governance Based on a Commons Perspective

3.1. Equity

Any public good is subject to some sort of distribution of rights and obligations amongst its stakeholders. Who does the work related to growth and maintenance, and who gets to benefit from what the resource has to offer? Is the resource available to those who need it, where and when they need it? Whose voices are heard when decisions are made that affect larger parts of the stakeholder community? Indeed, the question of who is considered to have a legit stake is in itself an important question in the context of a public good resource. While basic democratic values are obviously relevant, these issues are also pivotal in ensuring a wide commitment when few sanctions are available to force adherence to a common governance regime. The knowledge produced in the Myriad Laboratories is built on – and thus inherently part of – the global body of scientific knowledge about human genetic variations and their clinical significance. When they refuse to share their data that can contribute to further advancing the knowledge, they breach the commons´ logic. The community's response to Myriads enclosure policy is a reaction to the perceived inequity in the distribution of benefits (enclosure means that one key actor can reap disproportionally more benefits) and incurred costs (the generation of valuable data today is built upon the accumulated efforts and resources invested for many decades around the world).

3.2. Efficiency

For a governance arrangement to reach widespread support, it must provide an acceptable level of resource access and quality, at an acceptable cost, to members of the community. An inefficient arrangement will inevitably create a sense of a non-working system that might undermine its legitimacy and subsequent support. A particular challenging aspect of this is aligning, or balancing potentially diverse – and sometimes contradictory – values and requirements amongst different stakeholder groups. For instance, while uncertainty related to the clinical significance of a genetic variant is problematic for clinicians, it might represent a potential research question for the researchers. An efficient governance regime will have to make sure that such issues are reflected in the knowledge resource and its representations when possible.

3.3. Sustainability

Though a knowledge resource cannot be depleted, its long-term trajectory is normally subject to considerable dynamics. For a relatively novel field such as that of human genetic variants, the creation and addition of new knowledge is a main part of the dynamics. The evolution of knowledge and technology mutually drive each other. How can access, quality and costs be maintained within acceptable limits in this context of rapid change? In what direction is the trajectory moving and what are the issues driving and hampering its development? As new knowledge is primarily built on existing knowledge, an enclosure policy represents a potential threat to sustainability. Single actors growing too dominant can also undermine the community's trust and commitment, which in turn can jeopardise a responsible long term development. In the field of genetic knowledge, the continuous addition of new actors and their contributions and requirements is also an aspect that must be catered for in a long term perspective.

4. Concluding Remarks

The commons perspective advocates governance arrangements that favour the community as a whole while still allowing single stakeholders or groups of stakeholders to pursue their own interests. The ever increasing complexity related to the formation, maintenance and propagation of genetic and other biomedical knowledge will inevitably force the emergence of such arrangements. Fulk et al. explore the role of connective and communal public goods in discretionary repositories [8]. Information is discussed there as a hybrid (neither private nor public) good, where public benefit is achieved by individuals or companies acting out of their private interests. In the case discussed here the balancing of private interests is still in flux and this creates tensions for sustaining the discretionary repositories. It is our contention that a level of abstraction, where the various stakeholders can reach a common understanding and a shared goal, is a prerequisite for establishing a working governance regime. Based on the commons perspective, we have suggested *equity*, *efficiency* and *sustainability* as such an abstraction. As key principles, they may serve as a platform for translations and operationalisations into governance arrangements that will work for the community they are supposed to serve.

Acknowledgements

The authors would like to thank Morten Christoph Eike for fruitful discussions and for commenting on the manuscript.

References

[1] C. Bennett et al., A toolkit for incorporating genetics into mainstream medical services: Learning from service development pilots in England, *BMC health services research* **10.1** (2010), 125.
[2] H. Skirton, C. Patch and J. Williams, *Applied genetics in healthcare: A handbook for specialist practitioners*, Garland Science, 2005.
[3] R. Snyderman, Personalized health care: From theory to practice, *Biotechnology Journal* **7.8** (2012), 973-79.
[4] D. Gomez-Cabrero et al., Data integration in the era of omics: current and future challenges, *BMC systems biology* **8.2** (2014), I1.
[5] M. Shabani, B.M. Knoppers and P. Borry, From the principles of genomic data sharing to the practices of data access committees, *EMBO molecular medicine* **7.5** (2015), 507-509.
[6] E. Ostrom, *Governing the Commons: The Evolution of Institutions for Collective Action,* Cambridge University Press, Cambridge UK, 1990.
[7] E. Ostrom, J. Walker and R. Gardner, *Rules, games, and common-pool resources*, University of Michigan Press, Michigan US, 1994.
[8] J. Fulk, et al., Connective and Communal Public Goods in Interactive Communication Systems, *Communication Theory* **6.1** (1996), 60-87.
[9] M.D. McGinnis, An introduction to IAD and the language of the Ostrom workshop: a simple guide to a complex framework, *Policy Studies Journal* **39.1** (2011), 169-183.
[10] C. Hess and E. Ostrom, A framework for analyzing the knowledge commons, in C. Hess and E. Ostrom (eds.), *Understanding Knowledge as a Commons: from Theory to Practice*, MIT Press, Massachusetts US, 2006.
[11] J. Boyle, The second enclosure movement and the construction of the public domain, *Law and contemporary problems* **66.1/2** (2003), 33-74.
[12] S. Nguyen and S.F. Terry, Free the data: the end of genetic data as trade secrets, *Genetic testing and molecular biomarkers* **17.8** (2013), 579-80.
[13] G. Kolata, DNA Project Aims to Make Public a Company's Data on Cancer Genes, *The New York Times* (April 12 2013). <http://www.nytimes.com/2013/04/13/health/dna-project-aimsto-make-companys-data-public.html?pagewanted=all&_r=0>, accessed 15 September 2014.
[14] R.K. Yin, Case study research design and methods third edition. *Applied social research methods series* **5** (2003).
[15] G. Walsham, Doing interpretive research. *European journal of information systems* **15.3** (2006), 320-330.

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-303

Business Rules to Improve Secondary Data Use of Electronic Healthcare Systems

Jonathan BLAISURE[a,b,1] and Werner CEUSTERS [a,b]
[a] *Department of Biomedical Informatics, University at Buffalo, Buffalo NY, USA*
[b] *Institute for Healthcare Informatics, University at Buffalo, Buffalo NY, USA*

Abstract. The 'fit for purpose' paradigm used for data quality assessment in electronic healthcare record (EHR) systems is not so fit when assessed in the light of secondary data use. An analysis of the difficulties encountered in trying to use existing EHR data for cohort identification for prospective clinical trials and retrograde data analytics, revealed the root causes to fall in three categories: (1) issues in workflow and data registration, (2) preventable inadequacies in software configuration and personalization and (3) software development issues on the side of the vendor. By reviewing secondary data use requirements and formulating value adding business rules, development and data collection practices can be steered towards greater value in secondary data consumption.

Keywords. EHR, Secondary Use, Master Data Management, Business rules

1. Introduction

The quality of a healthcare system depends largely on the quality of data it relies upon to design, deliver, monitor and improve healthcare services that meet the needs of the targeted population. Though many definitions for data quality have been proposed, they predominantly express data quality as '*the totality of features and characteristics of a data set that bear on its ability to satisfy the needs that result from the intended use of the data*' [1] or, in summary, that datasets need to be '*fit for purpose*' [2]. Differences in definitions are then found in the amount and type of features and characteristics that relate to data quality and how they are to be measured in terms of various dimensions [3]. It is also this perspective on data quality that is typically adhered to in the healthcare domain where it forms the basis for data governance and data quality policies and procedures [4] and the translation thereof in business rules, i.e. '*statements that aims to influence or guide business processes in the organization*' [5] with the goal to ensure that the data are fit for purpose within the operational environment.

This 'fit for purpose qua intended use' perspective on datasets is in part what drives the design and use of electronic information systems within specific types of care settings such as inpatient charts, outpatient charts and practice management systems. Practice management systems deal mostly with scheduling, resource management and reimbursement. They can be standalone or integrated as part of an electronic health record (EHR). They contain operating data which strongly overlap other domain datasets

[1] Jonathan Blaisure, PhD student, Department of Biomedical Informatics, Jacobs School of Medicine and Biomedical Sciences, University at Buffalo, 77 Goodell street, Suite 540, Buffalo NY 14203, USA; E-mail: jcblaisu@buffalo.edu.

such as patient and provider demographics, schedules and reimbursement details. Outpatient systems are designed for ambulatory health care settings characterized by fast patient turnaround. They keep data for billing and auditing as well as snapshots of a patient's health to support treatment and care decision making. Inpatient systems deal with more severe cases, end of life, chronic illnesses and other long term diseases and injuries. These systems deal with data about the care delivered to patients and help manage patient populations in the facility itself.

This 'fit for purpose' quality perspective on datasets managed by individual information systems results, unfortunately, from a rather narrow interpretation of what the 'intended use' of the dataset exactly is: it narrows it down to the needs of the users of these specific systems. A problem then arises when data are to be re-used or integrated in other IS, for instance to monitor public health or improve global health services and policies. This is because individual information systems, when used within similar or overlapping geographic regions, exhibit overlap in patient population and information, but in different levels of detail, such level being determined by what makes the data fit for purpose for the users of each specific system in question, and this independent from whether this level fits the purpose of the users of the other systems it is communicating with or of the system(s) which integrate the data. The inpatient EHR of one organization may, for instance, record a patient being admitted for an asthma exacerbation that required overnight intervention. Some details about that treatment will be reported to the practice management system and finally some of those details will travel to the insurance company's record system where it is processed and eventually used to decide on payment. The patient may afterwards report for a follow-up exam by their primary care provider where more data about the same incident will be collected in the outpatient EHR system. Pooling data from these systems is very complex due to proprietary and badly documented designs that are not always in line with best practices for data interoperability [6]. In the scenario sketched above, each system may have some documentation about the patient event. Data gathered from each of these systems may support or conflict with one another or be such that from the data alone it cannot even be assessed that they report on the very same event. These conflicts require reconciliation or the documentation thereof.

An ideal situation would arise when electronic information systems would not just be designed to manage datasets that exhibit a level of detail that is fit for purpose for their own operational environment, but also for any secondary use thereafter. This does not mean that said systems would need to collect more data than required for the benefits of their own users – unless, of course, there would not be an additional burden – but rather that the data points would be more precise about what they describe. Our hypothesis is that the data curation problems we are experiencing today can be partly solved by implementing appropriate business rules. The work described in this paper demonstrates the feasibility of this proposal.

2. Methods

The University at Buffalo's Institute for Healthcare Informatics (IHI) primary mission is to gather fully identified healthcare data sets into a centralized secure environment where the data can be studied, documented and appropriately distributed for secondary data use projects. To date, the IHI houses a fully identified outpatient EMR database that contains about 650 thousand patient records (outpatient data) and a data set from a local healthcare

insurance provider (claims data) that contains about 1.2 million unique patients. The data is held under an institutional review board (IRB) protocol and data use agreements are signed by the data providers. The goal is to lower the barriers around data requests for secondary data usage since providers often do not have the resources or proper incentives to deliver data for research [7].

For the work presented here, we analyzed the last 19 secondary use data requests received (4 requests for claims data, 13 for cohort identification and 2 for data collection) to obtain more insight in the type of problems IHI staff encountered to deliver the data at the level of precision and quality expected by the researchers. A qualitative comparative analysis [8] was performed on the document trail (emails, meeting summaries, …) that resulted from the entire process consisting of (1) the initial – typically vague – data request, (2) the discussions between IHI staff and requesters to fully understand the data needs, (3) the construction of SQL queries, (4) the root cause analysis for requests that were not satisfiable, and (5) the development of business rules that could remediate the issues.

3. Results

We have identified three levels at which information gathering for secondary data use can be improved by implementing more appropriate business rules:

1. *Personnel workflow* – improving workflows that are (1a) ill-defined or contradictory to well-accepted data gathering practices that lead to data inconsistency at the practice management level or (1b) can be optimized for secondary use purposes.
2. *Software configuration* – the data collection software has certain configurations that can be modified by the system administrators. For example, a certain field in a screen that a healthcare professional uses during a patient visit can be marked as required rather than optional.
3. *Software development* – Identifying and documenting disparities between production and secondary data use requirements as a mechanism to present feedback to software vendors to improve future software releases.

4. Discussion

We provide here two examples of the analysis procedures applied and the business rules that were generated therefrom. Request A (Table 1) originally asked for a cohort of 'young adult survivors of cancer'. 'Young adult' translated after clarification in criterion C1, i.e. patients with ages 15 to 39, of which satisfaction could be determined based on the patient's date of birth in the EHR. Whereas what would count as 'cancer' was determined on the basis of ICD-codes, it was needed to relax 'young adult having cancer' (criterion C2) into 'being diagnosed with cancer' as it turned out not to be possible in every single case to determine on the basis of the EHR data (1) when the diagnosis was made, nor (2) when the cancer actually occurred since EHRs typically do not distinguish diagnoses from what they are about, i.e. the disease in the patient [9]. A date is always added when the diagnosis was recorded in the patient chart, but that can be a long time after the facts, for instance as the result of an anamnesis about prior disorders.

Table 1. Unsatisfiable criteria of request (A) for 'young adult cancer survivors'

Business Rule Level	/ Criteria # / Category
a. Workflow policy to require onset date of a disease recorded in the diagnosis in the provided onset field.	/ **C2** / **1&2**
b. Requiring an onset date field through the application configuration.	
c. Software development to incorporate other data sources to supplement current data.	/ **C3** / **3**

Table 2. Unsatisfiable criteria of request (B) for patients over 18 with childhood onset multiple sclerosis, an unaffected primary family member and living parents with available healthcare record.

Business Rule Level	/ Criteria # / Category
d. Workflow policy to encourage better family history record taking and use existing discrete data fields to record the data.	**C2** / **1, 2, 3**
e. Configure new fields in the application to capture data in discrete fields instead of free text.	
f. Require functionality in the application to appropriately implement the linking of family members.	
g. Workflow policy to encourage better family history record taking and use existing discrete data fields	**C4** / **1,3**
h. Require functionality in the application to enter information about relatives.	

This can be prevented by modifying workflows when entering a diagnosis by requiring a field to denote the disease onset date. This may not be known so other entries besides the onset date may be entered in the case of a fuzzy temporal value. E.g., the patient may not know exactly when the disease manifested but may know that it was during childhood. The last criterion, 'having survived cancer' (C3) presents another problem because healthcare facilities are not always aware of a patient's death, unless it happened on the premises. To illustrate this, only 1% of patients are marked as deceased and another 1% that are over 100 years of age and marked as alive. Other sources of data can be leveraged to fill in this gap such as the social security death index database. Request B (

Table 2 was a cohort search for 50 patients with (C1) childhood onset multiple sclerosis, (C2) an unaffected primary family member, (C3) current age greater than 18 and (C4) living parents with available healthcare record. The purpose of the study behind the request was to determine differences in environmental variables, phenotypical variables and genotype variations between the unaffected subjects and the affected subjects. C1 presents a similar problem as C2 discussed above. Information could in this case, however, be found in free text fields of the provider notes. These onset dates were located in different sections of the note and complicated techniques were required to extract the information. C2 could not be resolved since patient/relatives relationships are either poorly or not at all documented in the EHR. The EHR system from where the data had to be obtained does allow information to be entered about primary, secondary and tertiary relations but does not have a field to link genealogy from patient to patient. C3 could straightforwardly be calculated from the birth date of the patient. Although the source EHR system has fields in the family history section to record whether a patient's parents are living or their cause of death these fields are rarely used. C4 could thus not be satisfied. Business rule management is important, yet the literature on it is not as abundant as expected.[5]. Steinke classifies business rule types differently than we did: definition, guideline, mandate, and inference [10]. But he also takes the stance that reality should define the business rules used to govern the data. His approach is harmonious with ours, yet a full evaluation of the reality surrounding the information collections systems is costly and time-consuming. Recognizing the needs of secondary data users in creating business rules to oversee data collection presents a more practical solution.

5. Conclusion

This study was conducted in the context of one EHR system, though used by over 40 practices dispersed over the wider Buffalo area. Our analysis confirms earlier studies that the potentiality for secondary data use is not enough considered when EHR systems are designed, implemented and used [11]. Data required to perform certain analyses for cohort identification were found not to be easily obtainable due to (1) the insufficient facilities offered by the data collection systems to enter, store and retrieve required information and (2) the inadequate use of available facilities due to separate workflows and business practices across providers. The analysis of the secondary use data requirements made it nevertheless possible to develop business rules that help govern data entry and quality of data partially. Ironically, several of the requests for secondary use were issued by practitioners using the EHR system, and thus in part responsible for the lack of granularity in the recorded data. It remains to be investigated whether this awareness will lead to a smooth acceptation of and adherence to the proposed business rules and, as these processes become implemented and validated, will drive software vendors to update their systems to allow data to be entered in a way that is more faithful to reality.

Acknowledgement

This work was supported in part by Clinical and Translational Science Award NIH 1 UL1 TR001412-01 from the National Institutes of Health.

References

[1] Arts, D., d.N.F. Keizer, and G.J. Scheffer, *Defining and improving data quality in medical registries: a literature review, case study, and generic framework.* Journal of the American Medical Informatics Association, 2002. **9**(6): p. 600-611.
[2] Wang, R.Y. and D.M. Strong, *Beyond accuracy: What data quality means to data consumers.* Journal of Management Information Systems, 1996. **12**(4): p. 5.
[3] Sidi, F., et al. *Data quality: A survey of data quality dimensions*. IEEE.
[4] Davoudi, S., et al., *Data Quality Management Model (Updated).* Journal of AHIMA / American Health Information Management Association, 2015. **86**(10): p. 62-65.
[5] Shouhong, W. and W. Hai, *Business Rule Management for Enterprise Information Systems.* Information Resources Management Journal (IRMJ), 2010. **23**(1): p. 53-73.
[6] Kalra, D., et al., *ARGOS policy brief on semantic interoperability.* Studies in health technology and informatics, 2011. **170**: p. 1-15.
[7] Gilbert, R., H. Goldstein, and H. Hemingway, *The market in healthcare data.* BMJ-BRITISH MEDICAL JOURNAL, 2015. **351**: p. h5897.
[8] Kane, H., et al., *Using qualitative comparative analysis to understand and quantify translation and implementation.* Transl Behav Med, 2014. **4**(2): p. 201-8.
[9] Scheuermann, R.H., W. Ceusters, and B. Smith, *Toward an ontological treatment of disease and diagnosis.* Summit on Translat Bioinforma, 2009. **2009**: p. 116-20.
[10] Steinke, G. and C. Nickolette, *Business rules as the basis of an organization's information systems.* Industrial Management and Data Systems, 2003. **103**(1-2): p. 52-63.
[11] Hersh, W.R., et al., *Caveats for the use of operational electronic health record data in comparative effectiveness research.* Med Care, 2013. **51**(8 Suppl 3): p. S30-7.

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-308

Facilitators and Barriers of Electronic Health Record Patient Portal Adoption by Older Adults: A Literature Study

Gaby Anne WILDENBOS[ab1], Linda PEUTE[ab], Monique JASPERS[ab]

[a]*Center for Human Factors Engineering of Health Information Technology, Department of Medical Informatics, The Netherlands*
[b]*Academic Medical Center, University of Amsterdam, Department: Clinical Epidemiology, Biostatistics and Bioinformatics, Amsterdam Public Health research institute, Amsterdam, The Netherlands*

Abstract. Patient portal usage by older adults, patients aged 50 years old and above, is intended to improve their access and quality of care. Acceptance of patient portals by this target group is low. This paper discusses the results of a literature review to determine the facilitators and barriers that drive or inhibit older patients to adopt patient portals. Articles were included when they described an acceptance, adoption or usability evaluation study of a patient portal. From a total of 245 potentially relevant articles, 8 articles were finally included. We used the Unified Theory of Acceptance and Use of Technology (UTAUT) as a classification model to analyze factors influencing older adults' acceptance of patient portals. Main facilitators for acceptance were 'performance expectancy' and 'voluntariness of use' related to a higher level of education and experienced health. Main barriers were limited health literacy and motivation related to involuntariness to use a patient portal. Poor facilitation conditions (limited technology access and no prior knowledge on existence of a patient portal) hampered access to a portal. More thorough insight into the latter is needed to improve the reach and effectiveness of patient portals among older patients.

Keywords. Patient portal, older adults, elderly, acceptance, adoption, UTAUT

1. Introduction

Usage of patient portals is intended to improve access and quality of care and may result in improved health outcomes, especially for chronically ill patients [1]. Effective use of patient portals by older adults, patients aged 50 years old and above, is also expected to lower healthcare expenditures. Despite the associations of patient portal use with these favorable outcomes, voluntary uptake and use of patient portals by this target group have been low [2]. Older adults' *enrollment* to patient portals is expected to rise, since the elderly population is one of the fastest growing user segment of internet [2]. However, their *adoption* of patient portals is questionable. Besides a patient's age, aspects as health literacy level and socioeconomic status influence patient portal acceptance. Adoption of

[1] Corresponding author: Gaby Anne Wildenbos, J1B-109, Meibergdreef 9, 1105 AZ Amsterdam, The Netherlands; E-mail: g.a.wildenbos@amc.uva.nl

patient portals of older adults is further jeopardized by biological, psychological and social aging processes; it is thus relevant to examine which factors contribute to or impede the adoption of patient portals by this target group [3]. This is particularly relevant for chronically ill older adult patients, since it is foreseen that portals will increasingly fulfill a role in patients' self-management of their disease and care [4]. If patient portals indeed become a key tool for self-management and adoption rates of patient portals by older adults remain low, chronically ill older adult patients might be withheld from patient portal usage benefits and they might encounter difficulties due to a lack of non-patient portal services, such as a dedicated nurse practitioner to answer questions by telephone. It is thus needed to evaluate older adults' experiences and preferences for using a patient portal. The goal of this literature review is to determine the facilitators and barriers that drive older adult patients to adopt EHR patient portals by using the Unified Theory of Acceptance and Use of Technology (UTAUT), a prominent technology acceptance framework [5], as a classification model for those factors.

2. Methods

The database PUBMED was searched on the 7th of October 2016 for studies reporting on possible factors influencing the usage of patient portals. We used the following search strategy: "patients"[MeSH] OR "patients"[All Fields] OR "patient"[All Fields]) AND ("Electronic Health Records"[MeSH] OR "Telemedicine"[MeSH]) AND (factors[All Fields] OR barriers [All Fields] OR reasons [All Fields]) AND acceptance[All Fields]). We performed three additional searches including 'patient portal', 'barriers' and 'acceptance' in all fields. We limited all searches to articles in English published between January 2010 and July 2016. Studies were included if the abstract described an acceptance, adoption or usability evaluation study of a patient portal (1) with patients over 50 years old (2) in a hospital or primary care setting (3). We excluded articles reporting on a general status update of patient portals, with physicians as study population, reporting solely on one functionality of a patient portal, assisted living technologies, or a disease specific eHealth or mHealth tool.

The methodological quality of each article was evaluated by means of the Mixed Methods Appraisal Tool (MMAT), used for systematic mixed studies reviews, resulting in a standardized quality score across diverse type of study designs, study populations and sampling [6]. Per study, the first and second author independently identified facilitators and barriers and mapped these onto UTAUT concepts. If needed, a new concept in the UTAUT was introduced.

3. Results

3.1. General Characteristics of Included Studies

The searches resulted in 245 potentially relevant articles. Deduplication and abstract screening resulted in the rejection of 213 articles. Full text versions were not available for 2 articles and 22 articles were excluded after reading the full text version. This resulted in the inclusion of 8 articles for further analysis. The interrater reliability score

was κ=0,65. Five studies followed a qualitative design, with one using the UTAUT framework to formulate questions for the interviews and questionnaire [7].

3.2. Facilitators and Barriers of Older Adults' Patient Portal Usage

Table 1 and 2 respectively show the reported facilitators and barriers related to UTAUT concepts. Concerning facilitators, non-occurring UTAUT concepts were: effort expectancy, gender and age. We introduced one UTAUT sub-concept for performance expectancy and five sub-concepts for voluntariness of use. Concerning barriers, non-occurring UTAUT concepts were: performance expectancy and gender. We introduced two UTAUT sub-concepts for facilitating conditions and three sub-concepts for voluntariness of use.

Table 1. UTAUT categories related to facilitators of older adults' patient portal usage

UTAUT ***Extended UTAUT***	**Reported Facilitators**	**Age cluster**	**[Ref #]** ***Quality %***
Performance expectancy	Options for digital archiving and analysis of one's own medical data	55-75+	[8], *33*
	Beneficial aspects of patient portal: self-health monitoring by patient	Mean 56 Mean 51	[7], *40* [9], *40*
	Reread medical information at home	Mean 64	[1], *13*
Performance expectancy: benefits patient/provider relationship	Patient participation: patient portal facilitates influence of patient on their disease management and treatment	Mean 64	[1], *13*
	Option for sending secure messages to healthcare team via patient portal	65-79 Mean 56	[10], *13* [7], *40*
	Patient portal is neutral medium for delivering difficult news or 'difficult to hear' advice	55-75+	[8], *33*
Social influence	Availability of other person to help with usage	Mean 64	[1], *13*
	Physician or someone else thought use of patient portal would be useful for patient	Mean 64	[1], *13*
Facilitating conditions	Use via mobile device (smartphone, tablet)	65-79	[10], *13*
	Comfortable in using the internet	45-64, 65+	[11], *40*
Experience	Regular use of internet	Mean 64	[1], *13*
Voluntariness of use (VoU)	Ambition to learn how to use patient portal	50-100	[12], *40*
VoU: level of education	Higher level of education related to higher probability of using of patient portal	65-79 55-75+ Mean 55	[10], *13* [8], *33* [13], *13*
VoU: health interest & status	Better health conditions related to higher use of patient portal	65-79	[10], *13*
	Patients regularly searching health information on internet more likely to use patient portal	Mean 64	[1], *13*
VoU: dissatisfied current care communication	Dissatisfaction concerning current care communication	Mean 64	[1], *13*
VoU: Satisfied current care communication	Positive experiences healthcare clinic (careful listening, easy information explanation)	45-64, 65+	[11], 40
VoU: Cultural background	Cultural background might influence of use: Caucasians more positive attitude toward use and less connectivity problems than black patients	Mean 51	[9], *40*

Table 2. UTAUT categories related to barriers of older adults' patient portal usage

UTAUT ***Extended UTAUT***	**Reported Barriers**	**Age cluster**	**[Ref #]** ***Quality%***
Effort expectancy	Use of patient portal would be too complicated for patient according to patient	Mean 51	[9], *40*
Social influence	Older adults not able to use internet by themselves, only browse internet with help of others	65-79	[10], *13*
Facilitating conditions (FC)	Limited technology and/or internet access	55-75+ Mean 51 50-100	[8], *33* [9], *40* [12], *40*
FC: implementation issues	Limited to no prior knowledge on existence of patient portal	Mean 64 Mean 51 50-100	[1], *13* [9], *40* [12], *40*
FC: concerns	Concerns about privacy or security issues of medical data in patient portal	Mean 56 Mean 51	[7], *40* [9], *40*
Age	70 years old and above limited to no to use of patient portal	mean 64 55-75+ 65-79	[1], *13* [8], *33* [10], *13*
Experience	Lack of or limited technology experience and proficiency	55-75+ mean 56	[8], *33* [7], *40*
VoU: Health literacy	Limited health literacy	Mean 55	[13], *13*
VoU: satisfied with current care communication	Satisfied with the status quo	50-100	[12], *40*
	Wish to preserve in-person aspects of existing patient-provider relationships (i.r.t. limited health literacy)	Mean 56 Mean 51	[7], *40* [9], *40*
	Prefer to leave disease management to physician	Mean 64	[1], *13*
VoU: motivation	Lack of motivation to use patient portal	Mean 51	[9], *40*

4. Discussion

This literature review gives insight into the facilitators and barriers on patient portal acceptance by older adult patients, analyzed and clustered by means of the UTAUT. 18 facilitators predominantly concerned the UTAUT concepts 'performance expectancy: 'benefits patient/provider relationship', 'voluntariness of use: level of 'education', 'health interest & status', '(dis)satisfied with current care communication' and 'cultural background'. Twelve barriers predominantly concerned the UTAUT concepts 'facilitating conditions: implementation issues' & 'concerns' and 'voluntariness of use: 'health literacy', 'satisfied with current care communication' and 'motivation'.

We introduced 11 sub-concepts to the UTAUT, 6 related to facilitators and 5 to barriers. The UTAUT provides a theoretical framework for analyzing users' acceptance of health technology, but needs to be supplemented with concepts reflecting barriers and facilitators of older patients influencing their acceptance of patient portals. Another review suggests three additional constructs to the UTAUT for analyzing older users' home telehealth services acceptance: 'Doctor's Opinion', 'Computer Anxiety' and 'Perceived Security' [3]. These constructs correlate with the sub-concepts we introduced for 'performance expectancy' reflecting the (changing) patient/provider relationship by introduction of a patient portal and 'voluntariness to use' reflecting motivational reasons to use or not use a patient portal – such as (dis)satisfaction with current care communication, health literacy level and level of education. Yet, we found additional barriers related to access of digital health services for older adults, such as limited technology and internet access and privacy concerns on the medical data in the portal. Though most trend studies report that internet access issues of older adults will vanish over time, relying on everyday technology or generic internet use rates of seniors to

estimate digital health use may be misleading [14]. Seniors have used digital health tools at low rates with only modest increases from 2011 to 2014; these tools are not reaching most seniors and their underuse is associated with socioeconomic disparities, raising concerns about their ability to improve quality, cost, and safety of seniors' health care [14]. With the role of patient portals concerning patients' self-management growing, it is important to gain more sight on the conditions that hamper or facilitate the reach and acceptance of these portals by older patients.

5. Conclusion

Patient portal use is promoted in healthcare but acceptance rates of older patients are low. Older patients' expectancy of performance of a portal and higher education levels facilitate acceptance. Whereas a lower health literacy level and being satisfied with the status quo relate to involuntariness to use a patient portal. Poor facilitation conditions, such as limited technology and internet access, hamper older adults' access to a patient portal. Future research should focus on conditions for engaging older patient populations in patient portal usage within the broader context of patient profiles, the patient/provider relationship, decision making, provision and self-management of care.

References

[1] Ronda, M.C., Dijkhorst-Oei, L.T., Rutten, G.E., *Reasons and barriers for using a patient portal: survey among patients with diabetes mellitus.* J Med Internet Res, 2014. 16(11): p. e263.
[2] Atreja, A., et al., One size does not fit all: using qualitative methods to inform the development of an Internet portal for multiple sclerosis patients. AMIA Annu Symp Proc, 2005: p. 16-20.
[3] Cimperman, M., Makovec Brencic, M., Trkman, P. *Analyzing older users' home telehealth services acceptance behavior-applying an Extended UTAUT model*. Int J Med Inform, 2016. 90: p. 22-31.
[4] Winkelman, W.J., Leonard, K.J., Rossos, P.G., Patient-perceived usefulness of online electronic medical records: employing grounded theory in the development of information and communication technologies for use by patients living with chronic illness. J Am Med Inform Assoc, 2005. 12(3): p. 306-14.
[5] Venkatesh V, M.M., Davis G., David F., User acceptance of information technology: toward a unified view. 2003, MIS Quarterly. p. 425-478.
[6] Pluye, P., et al., *A scoring system for appraising mixed methods research, and concomitantly appraising qualitative, quantitative and mixed methods primary studies in Mixed Studies Reviews.* Int J Nurs Stud, 2009. 46(4): p. 529-46.
[7] Tieu, L., et al., *Barriers and Facilitators to Online Portal Use Among Patients and Caregivers in a Safety Net Health Care System*: A Qualitative Study. J Med Internet Res, 2015. 17(12): p. e275.
[8] Latulipe, C., Gatto, A., Nguyen, H.T., *Design Considerations for Patient Portal Adoption by Low-Income, Older Adults*. Proc SIGCHI Conf Hum Factor Comput Syst., 2015.
[9] Goel, M.S., et al., *Patient reported barriers to enrolling in a patient portal.* J Am Med Inform Assoc, 2011. 18 Suppl 1: p. i8-12.
[10] Gordon, N.P., Hornbrook, M.C., *Differences in Access to and Preferences for Using Patient Portals and Other eHealth Technologies Based on Race, Ethnicity, and Age: A Database and Survey Study of Seniors in a Large Health Plan*. J Med Internet Res, 2016. 18(3): p. e50.
[11] Butler, J.M., et al., *Understanding adoption of a personal health record in rural health care clinics: revealing barriers and facilitators of adoption including attributions about potential patient portal users and self-reported characteristics of early adopting users.* AMIA Annu Symp Proc. 2013: p. 152-61.
[12] Mishuris, R.G., et al., *Barriers to patient portal access among veterans receiving home-based primary care: a qualitative study*. Health Expect, 2015. 18(6): p. 2296-305.
[13] Davis, S.E., et al., *Health Literacy, Education Levels, and Patient Portal Usage During Hospitalizations*. AMIA Annu Symp Proc, 2015. 2015: p. 1871-80.
[14] Levine, M.D., et. al. *Trends in Seniors' Use of Digital Health Technology in the United States, 2011-2014*. J Am Med Assoc., 2016. 316(5).

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-313

Preventing Unintended Disclosure of Personally Identifiable Data Following Anonymisation

Chris SMITH [a,1]
[a] *Leeds Institute of Health Sciences, University of Leeds, United Kingdom*

Abstract. Errors and anomalies during the capture and processing of health data have the potential to place personally identifiable values into attributes of a dataset that are expected to contain non-identifiable values. Anonymisation focuses on those attributes that have been judged to enable identification of individuals. Attributes that are judged to contain non-identifiable values are not considered, but may be included in datasets that are shared by organisations. Consequently, organisations are at risk of sharing datasets that unintendedly disclose personally identifiable values through these attributes. This would have ethical and legal implications for organisations and privacy implications for individuals whose personally identifiable values are disclosed. In this paper, we formulate the problem of unintended disclosure following anonymisation, describe the necessary steps to address this problem, and discuss some key challenges to applying these steps in practice.

Keywords. anonymisation, unintended disclosure, personally identifiable data, human judgement, privacy

1. Introduction

Personally identifiable data [14] is captured and processed by care providers to inform service provision (e.g. emergency treatment). Simultaneously, these organisations are increasingly incentivised or mandated to share datasets with other organisations, or to make data available publicly, for secondary uses (e.g. clinical research) [15].

Legislation and ethical guidance governs data sharing by organisations. Generic legislation such as the Data Protection Act [19] and Human Rights Act [20] applies in the UK. Specific legislation also exists for the health domain, including the National Health Service Act 2006 [21] and Health and Social Care Act [22] in the UK, and the Health Information Portability Act (HIPAA) [23] in the US. Ethical guidance from advisory groups such as Research Ethics Committees [9] may supplement legislation.

Anonymisation can be applied to datasets prior to sharing in order to comply with legislation and ethical guidance. Attributes in a dataset are classified by the extent to which they facilitate identification of a ``data subject'' [11]. Risk of identification [6] posed by values in specific sets of attributes is then quantified using methods such as k-anonymity [18], l-diversity [13] and β-likeness [4]. Values are transformed [8, 17] to reduce this risk below an acceptable threshold [10] for a specific context.

[1] Corresponding author, Chris Smith, Leeds Institute of Health Sciences, University of Leeds, Level 10, Worsley Building, Clarendon Way, Leeds, LS2 9NL; E-mail: c.j.smith@leeds.ac.uk.

Human judgements are used to classify attributes as: *direct identifiers*: judged to enable identification in isolation, *indirect identifiers*: judged to enable identification in conjunction with other attributes within or outside the dataset, or *non-identifiers*: judged to provide minimal risk of identification [14]. Attribute classification, along with considerations of computational tractability [3, 7] and data utility [2, 16], determine whether specific attributes are included in: i) the anonymisation process, and ii) the shared dataset.

Classification requires knowledge of how attribute values contribute to identification. Legal and ethical guidance documents [10, 23] provide pre-defined classifications of common attributes (e.g. *Surname*) to assist organisations. However, datasets to be shared may exhibit characteristics such as high-dimensionality [1], structural dynamism, and complex provenance. Consequently, classification decisions may be based on incomplete and/or imperfect knowledge regarding attributes and their values. Failure of human judgements when making these type of information security decisions [5, 25], particularly in the evaluation of risk and uncertainty [24], and the consequences in terms of security and privacy breaches [12] has been previously recognised.

We focus on a problem that may arise from errors and anomalies during the capture and processing of health data: *personally identifiable values being placed into attributes of a dataset whose values are expected to contain non-identifiable data*. For example, a unique patient identifier residing in an attribute that relates to the reason for a referral due to an erroneous processing step that transposes values between attributes, or an informal staff policy to increase the speed of internal processes. Such attributes may be judged to contain non-identifiable values and omitted from anonymisation, but then included in a shared dataset. Consequently, organisations are at risk of sharing datasets that unintendedly disclose personally identifiable values through these attributes.

2. Unintended Disclosure of Personally Identifiable Data

We model a dataset as a set of entries, $d \in D$, where each entry is a tuple, $(v_1, \dots, v_n)$, that is composed of values for a set of attributes, $\mathrm{A} = \{\mathrm{a}_1, \dots, \mathrm{a}_n\}$, that relate to an individual, and V_a represents the set of distinct values held by an attribute, $a \in A$.

Patient Number	Date	Source	Reason
1120000	2014-04-01	ABC123	Cancer

Figure 1. Example entry from a dataset relating to patient referrals

Attributes are classified into one of the following distinct subsets of A to drive the anonymisation process: direct identifiers (I), indirect identifiers (Q), and non-identifiers (N), such that $\mathrm{I} \cup \mathrm{Q} \cup \mathrm{N} = \mathrm{A}$. Direct identifiers, $\mathrm{a} \in \mathrm{I}$, are removed. Non-identifiers, $a \in N$, are not considered in the anonymisation process, but may included in a shared dataset. We assume that at least one attribute is classified as a non-identifier, such that $\#(\mathrm{N}) \geq 1$. Indirect identifiers, $\mathrm{a} \in \mathrm{Q}$, are considered in anonymisation. For simplicity, and without loss of generality, we consider anonymisation to be applied over all attributes in Q rather than a subset of these attributes. Risk of identification is quantified based on combinations of values for the indirect identifiers, $V^*_{a_1} \times \dots \times V^*_{a_m}, \forall a_i \in Q$ for each entry in the dataset.

Transformations are applied to values of indirect identifiers to reduce the risk of identification. Specific transformations are dependent on the syntax and semantics of

attribute values. We do not consider aggregation over entries as a transformation, such that each entry remains associated with a single individual following anonymisation. Organisations iterate over a process of risk quantification and transformation to produce a dataset where the: i) risk of identification is reduced below an acceptable threshold, ii) data utility is sufficient for the intended use(s).

PID	Date	Source	Reason
1	2014-04	Specialist Clinic	Cancer

Figure 2. Example anonymised entry from a dataset relating to patient referrals

Figure 2 illustrates how the entry in Figure 1 might be anonymised. *Patient No* has been classified as a direct identifier and replaced by a unique (non-personal) identifier: *PID*. *Reason* has been classified as a non-identifier and retained in the dataset without transformation. *Date* and *Source* have been classified as indirect identifiers and transformed using generalisation.

Unintended disclosure of personally identifiable data has the potential to arise from judgements regarding attribute classification that are based on an expectation of the set of values, $E(V_a)$, held by the attribute, a, rather than the actual set of values, V_a. For example, the expectation that *Reason* will contain values that only relate to the condition for which they have been referred. Inconsistency between the expected and actual values can lead to an attribute that poses an identification risk being classified as a non-identifier and omitted from the anonymisation process. Actual values that contain personally identifiable data, such as *Patient No*, may reside in attributes classified as non-identifiers.

PID	Date	Source	Reason
1	2014-04	Specialist Clinic	Cancer-1120000

Figure 3. Example anonymised entry from a dataset relating to patient referrals with unintended disclosure

Figure 3 illustrates the problem of unintended disclosure. *Patient No* been included in the *Reason* attribute. Due to the classification of the *Reason* attribute as a non-identifier - based on the expectation that any values held by the attribute posed a minimal risk to identification - the attribute has been included in the dataset without transformation but clearly poses an identification risk.

Validation at different processing stages using techniques such as regular expressions may fail to prevent such scenarios. Methods may not be sufficiently restrictive due to their focus on the syntax rather than semantics of values. Additionally, datasets may be composed of entries from different organisations, which are subject to heterogeneous policies regarding data quality. Validation of these aggregated datasets might be insufficiently restrictive due to assumptions about upstream policies, or due to additional constraints, such as computational tractability.

3.Preventing Unintended Disclosure of Personally Identifiable Data

Prevention of unintended disclosure requires attributes to be classified based on *verification* rather than *expectation* of values. Verification ensures that any value, $v \in V_a$, of any attribute, $a \in A$, within any entry is drawn from a pre-defined set of values, V_a^*, for which the absence of personally identifiable data can be demonstrated. Any entry in the dataset would then be drawn from a defined space: $V_{a_1}^* \times \ldots \times V_{a_n}^*, \forall a_i \in A$. Verification could be integrated into attribute classification as follows:

- For $a \in A$, a set of values would be defined, V_a^*, which are drawn from a vocabulary for which the semantics and implications for identification of individuals are known, e.g. $V^*_{\text{Reason}} = \{\text{reason: cancer}, \text{reason: asthma}, \ldots\}$.
- For $a \in A$, the set of actual values held by the attribute, V_a, across all tuples would be determined, e.g. $V_{\text{Reason}} = \{\text{"Cancer"}, \text{"Asthma"}, \ldots\}$.
- For $v \in V_a \cap V_a^*$, a transformation function, δ, would be defined, to map actual values, V_a, of an attribute to values in V^*_a. Values not mapped would be omitted, or replaced with a *null* value, e.g. $\delta(\text{"Cancer"}) \rightarrow \text{reason: cancer}$.
- Classification would partition the set of attributes, A, based on known semantics of the extent to which they enable the identification of individuals, e.g. $I = \{\text{Patient No}\}$, $Q = \{\text{Date}, \text{Source}\}$, and $N = \{\text{Reason}\}$.

Verification would not only assist in preventing unintended disclosure, it would also provide a robust basis on which the identification risk posed by the values of different attributes could be quantified. Computational tractability and data utility could still be retained by evaluating identification risk over a subset of the attributes, Q. However, verification would ensure that classification of an attribute as a non-identifier does not risk the disclosure of personally identifiable data.

Classification based on verification of their actual values rather than human judgement is particularly important given the high-dimensionality, structural dynamism and complex provenance of datasets now captured by organisations. Robust judgement in the presence of such factors is a significant challenge for humans, yet such judgements are likely to be required more frequently within organisations in the future.

4.Challenges

Prevention of unintended disclosure presents challenges in practice, which include:

- **Computational Overhead:** Verification of attribute values against pre-defined sets is computationally intensive - requiring potentially vast numbers of comparisons. Efficient methods are required to minimise the time and resources required
- **Structural Dynamism:** Attributes and sets of attribute values can be subject to change over time - requiring changes to the verification process to ensure that it remains effective.
- **Vocabularies:** Attribute values must be compared against a pre-defined set of values - requiring relevant vocabularies to exist for each attribute. Organisations may be required to author such vocabularies if an appropriate vocabulary does not pre-exist for a particular attribute.
- **Technical Expertise:** Mapping of actual values, V_a, to a pre-defined set of values from a specific vocabulary, V^*_a, may not not necessarily be one-to-one and processing of certain formats for values may not be readily automated - requiring human involvement and technical expertise.

Without effective and efficient solutions to these challenges, organisations must decide whether to: (1) avoid sharing of datasets, or (2) share anonymised datasets and acknowledge the risk of unintended disclosure. This decision would be largely influenced by the legislative and ethical frameworks to which the organisation is subject.

5.Conclusion

Unintended disclosure of personally identifiable data has regulatory implications and poses governance challenges for data controllers. To situate sharing on a sound legal and ethical foundation, work is required to address the challenges above through novel tools and methods, vocabularies and ontologies, and education regarding anonymisation.

References

[1] Charu C Aggarwal. On K-anonymity and the Curse of Dimensionality. In *Proceedings of the 31st International Conference on Very Large Data Bases*, pages 901–909, 2005.
[2] Justin Brickell and Vitaly Shmatikov. The cost of privacy: destruction of data-mining utility in anonymized data publishing. *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 70–78, 2008.
[3] Jiwon Byun, Yonglak Sohn, Elisa Bertino, and Ninghui Li. Secure anonymization for incremental datasets. *Secure Data Management*, 4165:48–63, 2006.
[4] Jianneng Cao and Panagiotis Karras. Publishing Microdata with a Robust Privacy Guarantee. *Proceedings of the VLDB Endowment*, 5(11):1388–1399, 2012.
[5] Lorrie Faith Cranor. A framework for reasoning about the human in the loop. In *Proceedings of the 1st Conference on Usability, Psychology, and Security (UPSEC'08)*, pages 1–15, 2008.
[6] Khaled El Emam, Elizabeth Jonker, Luk Arbuckle, and Bradley Malin. A systematic review of re-identification attacks on health data. *PLoS ONE*, 6(12), 2011.
[7] Gabriel Ghinita, Yufei Tao, and Panos Kalnis. On the anonymization of sparse high-dimensional data. In *Proceedings of the 24th IEEE International Conference on Data Engineering (ICDE'08)*, pages 715–724, 2008.
[8] Yeye He and Jeffrey F Naughton. Anonymization of set-valued data via top-down, local generalization. *Proceedings of the VLDB Endowment*, 2(1):934–945, 2009.
[9] Health Research Authority. Research Ethics Committees (RECs), 2015.
[10] Information Commissioner's Office. Anonymization: Managing data protection risk - Code of Practice, 2012.
[11] ISO/TC215 Secretariat. Health Informatics - Pseudonymization (ISO/TS 25237:2008). Technical report, British Standards, 2008.
[12] Divakaran Liginlal, Inkook Sim, and Lara Khansa. How significant is human error as a cause of privacy breaches? An empirical study and a framework for error management. *Computers and Security*, 28(3-4):215–228, 2009.
[13] Ashwin Machanavajjhala, Daniel Kifer, Johannes Gehrke, and Muthuramakrishnan Venkitasubramaniam. L-Diversity: Privacy Beyond K-Anonymity. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 1(1):1–52, 2007.
[14] Arvind Narayanan and Vitaly Shmatikov. Myths and fallacies of "personally identifiable information". *Communications of the ACM*, 53(6):24, 2010.
[15] Nuffield Council on Bioethics. The collection, linking and use of data in biomedical research and healthcare: ethical issues, 2015.
[16] Paul Ohm. Broken Promises of Privacy: Responding to the Surprising Failure of Anoymization. *UCLA Law Review*, 57:1701–1777, 2010.
[17] Pierangela Samarati and Latanya Sweeney. Protecting Privacy when Disclosing Information: k-Anonymity and Its Enforcement through Generalization and Suppression. Technical report, SRI Computer Science Laboratory, Palo Alto, CA, 1998.
[18] Latanya Sweeney. k-anonymity: A model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 10(05):557–570, 2002.
[19] UK Legislation. Data Protection Act, 1998.
[20] UK Legislation. Human Rights Act 1998, 1998.
[21] UK Legislation. National Health Service Act 2006, 2006.
[22] UK Legislation. Health and Social Care Act 2012, 2012.
[23] U.S. Department of Health and Human Services. The HIPAA Privacy Rule, 2000.
[24] Ryan West. The Psychology of Security: Why do good users make bad decisions. *Communications of the ACM*, 51(4):50–79, 2008.
[25] Charles Cresson Wood and William W. Banks. Human error: an overlooked but significant information security problem. *Computers & Security*, 12(1):51–60, 1993.

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-318

Protecting Privacy of Genomic Information

Jaime DELGADO[1], Silvia LLORENTE and Daniel NARO
Distributed Multimedia Applications Group (DMAG), Computer Architecture Dept. (DAC), Universitat Politècnica de Catalunya (UPC)

Abstract. The ISO/IEC committee in charge of standardizing the well-known MPEG audiovisual standards has launched, in cooperation with the ISO committee on Biotechnology, a new activity for efficient compressed storage and transmission of genomic information. The paper presents proposals for adding privacy and security to such in-progress standards.

Keywords. Privacy, Security, Genomic Information

1. Introduction

Nowadays, there is an increasing amount of genomic information being generated for different purposes: research, genetic analysis, precision medicine, search of relatives, forensics, etc. This means that the information generated will probably need to be stored, transferred and, ideally, protected against attacks.

On the one hand, several attacks have already been identified, usually to find out if an individual's genome is inside a set of genomes [1], but also to infer regions of the genome which have to remain private [2].

On the other hand, there is an initiative of the Moving Picture Experts Group (MPEG) Standardization Committee [3] for the representation of genomic information in a compressed way, including security and privacy aspects.

The authors have presented several proposals (such as [4]) as answer to the MPEG Committee call for proposals [5], but we concentrate on the one that includes the provision of security and privacy mechanisms to support the storage and transmission of genomic information.

In the next sections, we describe the MPEG Committee activities on genomic information, our proposal for security and privacy and some conclusions and future work.

2. Methods: The Moving Pictures Experts Group (MPEG) Work on Genomics

The Moving Picture Experts Group (MPEG) [3] is a working group of ISO/IEC (ISO/IEC JTC 1/SC 29/WG 11). Since 1988, the group has produced standards for coded representation of digital audio and video and related data.

[1] Corresponding author, Jaime Delgado, Universitat Politècnica de Catalunya, Jordi Girona, 1-3, 08034 Barcelona, Spain; E-mail: jaime.delgado@ac.upc.edu.

Based on their successful previous experience in audiovisual content compression, a new initiative inside MPEG to provide compression mechanisms for genomic information was started two years ago. After a detailed process of obtaining requirements, a call for proposals was launched in July 2016 [5] in order to provide solutions to the genomic information compression and representation problem. The 15 answers to this call have been discussed in the 116th meeting held in Chengdu in October 2016 [6].

The paper focuses on the proposal presented by a group of organizations and companies led by the authors [4]. The contribution responds to the transport requirements [7], defining a format for storage and transport of genomic information, including aspects like metadata, definition of security mechanisms and application of privacy rules.

3. Results: Privacy and Security in Genomics contributed to MPEG

In [4], we propose a format (that we call GENIFF, for GENomic Information File Format) based on ISOBMFF [8] to support the inclusion of compressed and uncompressed genomic information. Moreover, we also provide mechanisms to include metadata associated to the information stored inside the format. Metadata elements may apply to a study, an individual, partial or complete genomic information, etc.

Apart from generic metadata describing things like the genetic study done, the machine used for generating the genomic information and other kind of metadata defined by European Bioinformatics Institute (EBI) [9], we have focused on providing placeholders to apply security and privacy to the genomic information. In this way, only authorized users will be able to access to the information contained in the file. Security and privacy elements may apply to different levels, like the complete study or only to specific information, for example, a SAM / BAM file [10] or a Variant Call File (VCF) [11]. In this way, we can provide a high level of flexibility.

3.1. Privacy Provision, Use of XACML to Authorize Access to Genomic Information

In order to provide privacy when accessing genomic information, we use XACML [12] rules. First ideas on how to use XACML where already presented to the MPEG Committee in [13] and are reflected in our proposal for representing genomic information [4]. We are also using XACML rules to provide privacy protection when accessing medical information, as described in [14]. The rules for controlling access to genomic information may include, among other, the following information:

- who is able to access to the genomic information (user roles or individuals);
- what information can be accessed (the complete file, a chromosome, etc.);
- when it can be accessed;
- with which purpose (genetic analysis, anonymized study, etc.);
- if the data provider has to be informed when information is accessed;
- which permission is given (`viewFile`, `viewChromosome`, etc.).

The rules can be included inside our file format (GENIFF), or even inside the existing formats for genomic information, like SAM or BAM [10].

The inclusion of the rules inside the genomic files allows us to extract them and authorize access to the file according to the permissions defined in the rules. The workflow of operations required to authorize the access to genomic information is shown in Figure 1. It works as follows:

1. A user requests, to a repository, access to some specific genomic information. To do so, an access request is sent including, among other, the following context information: user or role requesting access, the operation requested, the time when the request is done, which information wants to be accessed and the level of granularity (i.e. complete file, chromosome, etc.).
2. The genomic information repository extracts the rule(s) from the requested genomic information and asks, to the Authorization point, for user authorization according to the rule(s) and the request received.
3. The authorization result could be Permit, Deny or Not Applicable.
4. In the case of Permit, the access is permitted, so the requested genomic information should be decrypted and given to the user from the repository.

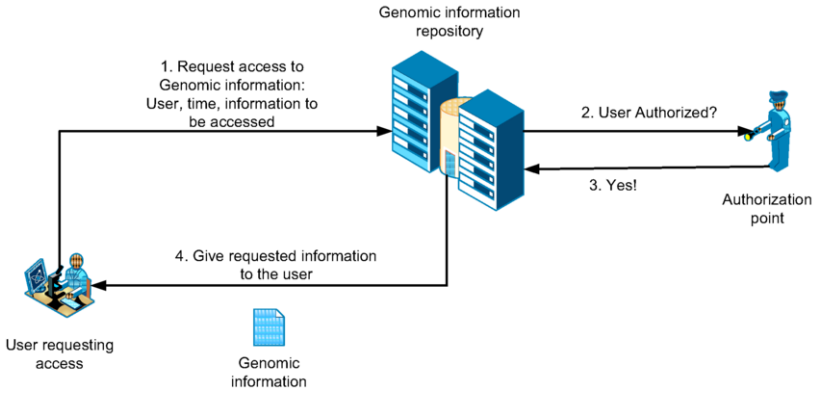


Figure 1. Genomic data authorization flow.

Figure 2 shows a snippet of an XACML rule giving VIEW permission to a physician for the file `genomic-file.sam`.

3.2. Encrypting Portions of the Data

When securing genomic information, one of the preferred strategies is to aggregate results of multiple individuals and to only respond to queries concerning the whole pool of data. Such queries might be, for example, the frequency of a certain mutation in the studied population. This is intended to ensure the privacy at the individual level, although this approach was proven to have some flaws [1].

The result of MPEG's work has to be a format for genomic information, where the encryption of the data is used as a way to enforce the privacy rules. Certain regions of the DNA can be considered as safe to make them public, while others require a key to decrypt the content. By encrypting those regions, we ensure that the file can be transmitted, and the access to the content can vary depending on the access the user asks for. The actual policy to decide which regions should be protected is left to the user creating the file, but one approach might be, as in [2], encrypting the regions to remain private, but also include those allowing inference attempts.

The encryption is done on well-defined regions of the DNA, for example for a given chromosome, using symmetric encryption. Homomorphic encryption might add the benefits to enable secure computations on encrypted content (e.g. [15]); however, this encryption methodology appears to be tightly bounded to the intended usage, thus limiting broad applicability. Therefore, our proposed format uses symmetric encryption, leveraging on the employed codec for better granularity (e.g. BAM's blocks).

```
<Rule RuleId="urn:oasis:names:tc:xacml:3.0:RuleSAM" Effect="Permit">
  <Description> A physician may view the genomic information file for which he or
    she is the designated primary care physician
  </Description>
  <Target>
    <AnyOf>
      <AllOf>
        <!-- Which kind of user: physician -->
        <Match MatchId="urn:oasis:names:tc:xacml:1.0:function:string-equal">
          <AttributeValue DataType="http://www.w3.org/2001/XMLSchema#string">
            physician
          </AttributeValue>
          <AttributeDesignator MustBePresent="false"
            Category="urn:oasis:names:tc:xacml:3.0:role" AttributeId="role"
            DataType="http://www.w3.org/2001/XMLSchema#string"/>
        </Match>
        <!-- Which resource -->
        <Match MatchId="urn:oasis:names:tc:xacml:1.0:function:regexp-string-match">
          <AttributeValue DataType="http://www.w3.org/2001/XMLSchema#string">
            genomic-file.sam
          </AttributeValue>
          <AttributeDesignator MustBePresent="false"
            Category="urn:oasis:names:tc:xacml:3.0:attribute-category:resource"
            AttributeId="urn:oasis:names:tc:xacml:1.0:resource:resource-id"
            DataType="http://www.w3.org/2001/XMLSchema#string"/>
        </Match>
        <!-- Which action  -->
        <Match MatchId="urn:oasis:names:tc:xacml:1.0:function:string-equal">
          <AttributeValue DataType="http://www.w3.org/2001/XMLSchema#string">
            VIEW
          </AttributeValue>
          <AttributeDesignator MustBePresent="false"
            Category="urn:oasis:names:tc:xacml:3.0:attribute-category:action"
            AttributeId="urn:oasis:names:tc:xacml:1.0:action:action-id"
            DataType="http://www.w3.org/2001/XMLSchema#string"/>
        </Match>
      </AllOf>
    </AnyOf>
  </Target>
</Rule>
```

Rule effect,

Authorized role:

Resource: genomic-

Permission

Figure 2. XACML rule giving VIEW permission to a SAM file.

4. Discussion, Conclusions and Future Work

We have developed software to demonstrate the feasibility of the format and its granular access control. We have implemented some policy rules with XACML and used real SAM/BAM files to evaluate the encryption of content for supporting the access control. The results are satisfactory for the moment.

Our next step is aligned with the standardization process. Some of the presented ideas have been initially accepted for consideration by the MPEG Committee. Now it is time to perform some formal tests to check their feasibility. For this purpose, a set of experiments has been defined [16] and the results have to be presented for discussion at the next MPEG meeting, in January 2017.

In particular, we will further analyze and implement the presented approach in experiment 4, which will check that the genomic information representation format defines the proper tools for allowing selective access to the content for both data storage and transfer. In addition, these tools should support access control (privacy and security) and high-level metadata applying to different genomic information contained in the format. We will particularly focus on providing tools for privacy, security and metadata definition inside the genomic information.

Acknowledgements

The work presented in this paper has been partially supported by the Spanish Government under the project: Secure Genomic Information Compression (GenCom, TEC2015-67774-C2-1-R).

References

[1] Nils Homer et al., *Resolving individuals contributing trace amounts of DNA to highly complex mixtures using high-density SNP genotyping microarrays*. PLoS genetics, 4(8):e1000167, 2008.
[2] Nyholt, Dale R and Yu, Chang-En and Visscher, Peter M. On Jim Watson's APOE status: genetic information is hard to hide. European journal of human genetics : EJHG march 2009.
[3] ISO/IEC JTC 1/SC 29/WG 11, *Moving Picture Experts Group (MPEG),* http://mpeg.chiariglione.org/.
[4] MPEG2016/M39175, *GENIFF (GENomic Information File Format), a proposal for a Secure Genomic Information Transport Layer (GITL) based on the ISO Base Media File Format*, Chengdu, October 2016.
[5] ISO/IEC JTC 1/SC 29/WG 11 - ISO/TC 276/WG 5 MPEG2016/N16320, *Joint Call for Proposals for Genomic Information Compression and Storage*, Geneva, June 2016.
[6] MPEG2016/M38890, *AHG on Requirements on Genome Compression and Storage*, Chengdu, October 2016.
[7] ISO/IEC JTC 1/SC 29/WG 11 - ISO/TC 276/WG 5 MPEG2016/N16323, *Requirements on Genomic Information Compression and Storage*, Geneva, June 2016.
[8] ISO/IEC IS 14496-12, *Information technology - Coding of audiovisual objects - Part 12: ISO base media file format*, Fifth edition, December 2015.
[9] European Bioinformatics Institute (EBI), http://www.ebi.ac.uk/, 2016.
[10] *Official Sequence Alignment/Map (SAM) Format Specification*, https://samtools.github.io/hts-specs/, 2016.
[11] *Official Variant Call Format (VCF) Format Specification*, https://samtools.github.io/hts-specs/, 2016.
[12] OASIS Standard, *eXtensible Access Control Markup Language (XACML) Version 3.0*, http://docs.oasis-open.org/xacml/3.0/xacml-3.0-core-spec-os-en.html, January 2013.
[13] Jaime Delgado, Silvia Llorente, MPEG2015/M36405 - *Some application scenarios for privacy and security requirements on genome usage, compression, transmission and storage*, Warsaw, June 2015.
[14] Jaime Delgado, Silvia Llorente, Martí Pàmies and Josep Vilalta, *Security and Privacy in a DACS*, Exploring Complexity in Health: An Interdisciplinary Systems Approach, doi: 10.3233/978-1-61499-678-1-122, IOS Press, 122-126, 2016.
[15] Jung Hee Cheon, Miran Kim, and Kristin Lauter, *Homomorphic computation of Edit Distance*, Pages 194-212 Springer Berlin Heidelberg 2015.
[16] ISO/IEC JTC 1/SC 29/WG 11- ISO/TC 276/WG 5 MPEG2016/ N16526, *Core Experiments on Genomic Information Representation*, Chengdu, October 2016.

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-323

Clinical Data Warehouse Watermarking: Impact on Syndromic Measure

Guillaume BOUZILLE[abcd,1] , Wei PAN[e], Javier FRANCO-CONTRERAS[ef], Marc CUGGIA[abcd], and Gouenou COATRIEUX[e]
[a]*INSERM, U1099, Rennes, F-35000, France*
[b]*Université de Rennes 1, LTSI, Rennes, F-35000, France*
[c]*CHU Rennes, CIC Inserm 1414, Rennes, F-35000, France*
[d]*CHU Rennes, Centre de Données Cliniques, Rennes, F-35000, France*
[e]*Institut Mines-Telecom;Telecom Bretagne; Latim Inserm UMR1101, Brest, France.*
[f]*WaToo, Brest, France*

Abstract. Watermarking appears as a promising tool for the traceability of shared medical databases as it allows hiding the traceability information into the database itself. However, it is necessary to ensure that the distortion resulting from this process does not hinder subsequent data analysis. In this paper, we present the preliminary results of a study on the impact of watermarking in the estimation of flu activities. These results show that flu epidemics periods can be estimated without significant perturbation even when considering a moderate watermark distortion.

Keywords. Traceability, Clinical data warehouse, Watermarking

1. Introduction

The widespread adoption of Electronic Health Records (EHRs) and the emergence of clinical data warehouses (CDWs) offer nowadays great opportunities to share and reuse patients' data for secondary purposes, such as medical research [1]. This implies using appropriate infrastructures that comply with the policies of privacy and data protection dedicated to patients' data. In this context, secondary use of health data requires a governance dedicated to regulate access to data and technologies ensuring reliable access matching this governance. A typical organization offering this kind of service is made of two parts: first, a CDW that embodies the infrastructure allowing to efficiently reuse data; a regulatory board that supervises ethical and legal aspects of studies that aim to reuse patient's data. The conditions to be fulfilled to access patient data for secondary purposes depend on countries' legislation but it remains on several unshakable criteria: parsimony, deidentification, authorization and traceability.

Nowadays, most CDWs technologies, such as i2b2, STRIDE or other custom systems are consistent with these requirements [2-4]. However, organizations are now in the way to share data in order to reach a new scale concerning sample size or

[1] Guillaume Bouzillé, LTSI, équipe-projet Données Massives en Santé - Université de Rennes 1, Campus de Villejean - Bâtiment 6, 35043 Rennes Cedex - France, ; E-mail: guillaume.bouzille@univ-rennes1.fr.

geographical coverage. Such exploitation requires new centralized or distributed large scale platform (e.g SHRINE or EHR4CR [5-6]) complying with the regulatory. In this context, new levels of requirements for data security and patient privacy have to be provided. Indeed, patient's datasets are likely to be exposed outside healthcare organizations increasing the risk for malicious data collection. One key security measure consists in ensuring the traceability of datasets and data processing. Watermarking is a promising approach allowing the embedding of a message containing traceability information (e.g., user identity, date of access) into the database by slightly modifying some of its attributes' values. Watermarking provides free access to the data while keeping it protected through the embedded message and it is complementary to other security mechanisms already deployed. Nevertheless, it is necessary to ensure that the distortion resulting from the watermark does not impair the interpretation or any subsequent data analysis [7].

In this paper, we evaluate a database watermarking method to ensure traceability data sharing in the context of epidemiologic trends analysis. More precisely, we evaluate the impact of watermarked data on the production of flu activity estimates.

2. Methods

2.1. Dataset to be Watermarked

We extracted from eHOP (the CDW of the academic hospital of Rennes [8]) all flu PCR tests that were performed on patients between January 1, 2011 and February 13, 2016. We considered all PCR tests carried out, regardless of whether the result was negative or positive. The aim was to get a signal connected with influenza-like illness (ILI) symptoms and not only with the flu. This datasource is known for having a high correlation with standard ILI estimates provided by the french Sentinel network, at a regional and national scale.

2.2. Watermarking Algorithm and Parameterization

In order to deploy this test, we implemented the scheme developed in [9], the advantage of which is that it injects a constant distortion. More clearly, this one embeds a sequence of bits into the values of an integer attribute A_i of dynamic range [*min*, *max*] by adding the quantity Δ to approximately a half of its values and $-\Delta$ to the others. The length of the message and its robustness to database alterations (i.e., the capability to retrieve the message) depend on Δ. The greater Δ, the more robust or the longer the message can be (for more details see [9]). This watermarking algorithm was applied on the date of PCR tests' realization with 28 different Δ, ranging from 1 to 28 days. For each Δ, we replicated the watermarking procedure 27 times to assess the variability of its induced distortion. We thus created 756 different watermarked datasets to be compared with the original one. The embedded message held 100 bits.

2.3. Statistical Analyses

For each dataset (original and watermarked ones), we computed weekly ILI activity incidences: every PCR date was modified to match the monday of the same week,

according to the ISO 8601 for representation of dates. We then counted PCR realisations according to their modified dates.

The evaluation of the distortion induced by the watermark procedure relied on two indicators. First we used the Pearson's correlation coefficient (PCC) to assess how watermarked data could produce ILI activity indexes associated with the original one. Second, we used the Normalized Root Mean Squared Error (NRMSE) to measure how ILI incidences computed with watermarked data deviate from the original ones. For each Δ, we performed a non-parametric bootstrap procedure with 1,000 replicates of the original sample of 27 watermarked datasets. We used the bootstrap replicates to compute 95% confidence intervals for PCC and NRMSE estimates.

To assess the effect of watermarking on the decision making process, we also applied the statistical model used in routine practice by the french sentinel network to detect influenza epidemic periods: the Serfling periodic regression model [10]. We assessed differences in epidemics periods detected with watermarked data and reference data. All analyses were performed on R (version 3.3.1).

3. Results

We extracted 10,555 PCR tests between January 1, 2011 and February 13, 2016, from the CDW of CHU-RENNES. They were performed on 2,965 different patients. The watermarking process applied on this dataset took approximately 2 minutes, a time that strongly depends on database server read/write performance. The entire set used in the study contained 756 watermarked datasets. The dates of reference data were distributed on 267 weeks, that is to say we produced 267 weekly ILI activity estimates. Watermarked data produced 267 to 271 weekly ILI activity estimates depending on the strength of the distortion parameter. We calculated PCCs and NRMSE on the period shared by the reference and watermarked data. PCC decreased from 0.99 (95% CI: 0.99; 0.99) to 0.64 (95% CI: 0.64 ; 0.64) for a time shift between 1 day and 28 days (Table. 1). NRMSE increased from 3.5 (95% CI: 3.5; 3.5) to 19.4 (95% CI: 19.3 ; 19.5) respectively for a Δ value of 1 day to 28 days (Table 1).

Table 1. Pearson's correlation coefficients and Normalized Root Mean Squared Errors for Δ=[1,7,14,21,28]

Δ (days)	Pearson's correlation coefficient (95% CI)	Normalized Root Mean Squared Error (95% CI)
1	0.99 [0.99;0.99]	3.5 [3.5; 3.5]
7	0.90 [0.89;0.90]	10.1 [10.1;10.2]
14	0.76 [0.76;0.76]	15.7 [15.6;15.8]
21	0.67 [0.67;0.67]	18.4 [18.4;18.5]
28	0.64 [0.64;0.64]	19.4 [19.3;19.5]

We used Serfling's periodic regression to detect epidemics. All bootstrap replicates with a Δ below 7 days allowed to detect the 6 epidemics periods, which were present in the original data. With Δ up to 8 days, all epidemics were not detected: a Δ of 8 days resulted in 9% of bootstrap replicates with 7 detected epidemics ; a Δ of 21 days resulted in a detection of 5 epidemics in 613 out of 1,000 replicates and a Δ of 28 days resulted in 5 epidemics detected in 99% of the replicates. For replicates which correctly detected the 6 epidemics, delays to detect epidemics compared to the true epidemics periods are depicted in Figure 1.

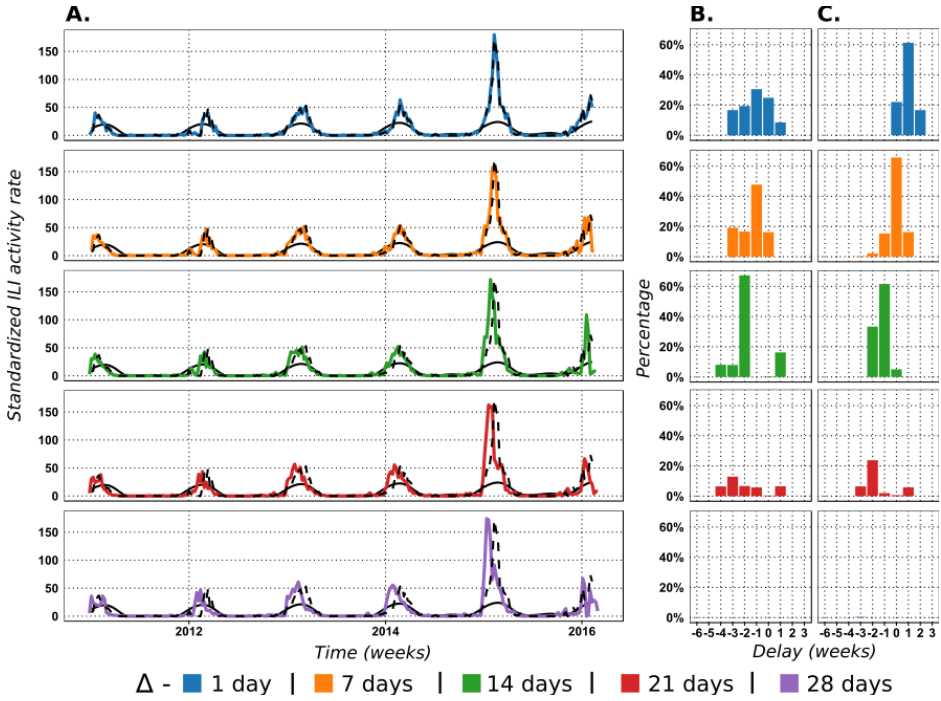


Figure 1. A - Weekly ILI activity estimates from watermarked data. Black dashed lines represent ILI activity computed from original data ; black solid lines are Serfling's model fitted on original data. **B and C -** Delays to detect respectively start and end of epidemics from watermarked data. Percentages are computed on bootstrap replicates which detected the 6 true epidemic periods.

4. Discussion

Watermarking of clinical data appears to be a fairly simple and fast process to ensure traceability of health big data. Traceability represents an additional security level that is currently rarely considered when reusing patients' data but of great concern with the expansion of health data sharing. We assessed a watermark of a datamart –stemming from our CDW– which was dedicated to influenza surveillance. We chose to watermark dates of PCR laboratory results. Indeed, watermarking dates seems to be a rational choice when producing time series of disease activity, because this field is mandatory to produce the time-dependent measures and therefore may not to be deleted. We showed that moderate distortion impacts the reliability of produced incidences in a way that will not drastically change their meaning. For instance, all flu epidemics periods were detected up to a distortion of 7 days, with slight enlargement of epidemics periods. However, with stronger distortion, several epidemics are not detected or present wrong periods. This suggests that watermark procedures have to be carefully tuned so as to preserve the quality of data analyses. Syndromic surveillance also implies forecasting, which can rely on predictive algorithms. It seems interesting to conduct further work to assess how watermarking disrupt learning process, for instance to forecast disease activity.

Watermarking is of particular concern for syndromic surveillance, which requires sharing data from different sources distributed on a geographic area, before to produce aggregated data. This necessarily implies exposing data outside infrastructures generating data, which typically are CDWs. However, CDWs are dedicated to store data for secondary use with multi purposes. Yet, it seems hard to perform a watermark of a whole CDW, which could be compatible to all use cases for which they are intended. As a consequence, we believe that, in the context of health data traceability, watermarking must be performed at the final step of the data sharing process, when the data are exported for a given use case. Watermarking parameters can be tuned according to the use case, to ensure distortion which will be acceptable, without jeopardizing other use cases of CDWs.

Acknowledgements

This work was supported in part by the French National Research Agency-Project ANR inside the INSHARE (INtegrating and Sharing Health dAta for Research) project (grant no. ANR-15-CE19-0024).

References

[1] Cohen B, Vawdrey DK, Liu J, Caplan D, Furuya EY, Mis FW, et al. Challenges Associated With Using Large Data Sets for Quality Assessment and Research in Clinical Settings. *Policy Polit Nurs Pract*. (2015) 16(0):117–24.

[2] Murphy SN, Weber G, Mendis M, Gainer V, Chueh HC, Churchill S, et al. Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2). *J Am Med Inform Assoc*. (2010);17(2):124–30.

[3] Lowe HJ, Ferris TA, Hernandez PM, Weber SC. STRIDE – An Integrated Standards-Based Translational Research Informatics Platform. *AMIA Annu Symp Proc*. 2009;2009:391–5.

[4] Holmes JH, Elliott TE, Brown JS, Raebel MA, Davidson A, Nelson AF, Chung A, La Chance P, Steiner JF, "Clinical research data warehouse governance for distributed research networks in the USA: a systematic review of the literature.", *J Amer Med Inf Assoc*. 21(4), (2014), 730-736

[5] Weber GM, Murphy SN, McMurry AJ, MacFadden D, Nigrin DJ, Churchill S, et al. The Shared Health Research Information Network (SHRINE): A Prototype Federated Query Tool for Clinical Data Repositories. *J Amer Med Inf Assoc*. (2009) 16(5):624–30..

[6] De Moor G, Sundgren M, Kalra D, Schmidt A, Dugas M, Claerhout B, et al. Using electronic health records for clinical research: the case of the EHR4CR project. *J Biomed Inform*. (2015) 53:162–73.

[7] Franco-Contreras J, Coatrieux G. Robust Watermarking of Relational Databases With Ontology-Guided Distortion Control. *IEEE Trans. Inf. Forens. Sec*. 2015 Sep;10(9):1939–52.

[8] Delamarre D, Bouzille G, Dalleau K, Courtel D, Cuggia M. Semantic integration of medication data into the EHOP Clinical Data Warehouse. *Stud Health Technol Inform*. 2015;210:702–6.

[9] Franco-Contreras J, Coatrieux G, Cuppens F, Cuppens-Boulahia N, Roux C. Robust Lossless Watermarking of Relational Databases Based on Circular Histogram Modulation. *IEEE Trans. Inf. Forens. Sec*.. 2014 **9**(3):397–410.

[10] Serfling RE. Methods for current statistical analysis of excess pneumonia-influenza deaths. *Public Health Rep*. 1963 **78**(6):494–506.

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-328

Security Policy and Infrastructure in the Context of a Multi-Centeric Information System Dedicated to Autism Spectrum Disorder

Mohamed BEN SAID [a,1], Laurence ROBEL [b], Bernard GOLSE [b], Jean Philippe JAIS [a]
[a] *Paris Descartes University, Faculty of Medicine, Department of Biostatistics and Medical Informatics, APHP - Necker Enfants Malades Hospital, Paris, France*
[b] *Department of Child Psychiatry, APHP - Necker Enfants Malades Hospital, Paris, France*

Abstract. Autism spectrum disorders (ASD) are complex neuro-developmental disorders affecting children in their early age. The diagnosis of ASD relies on multidisciplinary investigations, in psychiatry, neurology, genetics, electrophysiology, neuro-imagery, audiology and ophthalmology. In order to support clinicians, researchers and public health decision makers, we designed an information system dedicated to ASD, called TEDIS. TEDIS was designed to manage systematic, exhaustive and continuous multi-centric patient data collection via secured Internet connections. In this paper, we present the security policy and security infrastructure we developed to protect ASD' patients' clinical data and patients' privacy. We tested our system on 359 ASD patient records in a local secured intranet environment and showed that the security system is functional, with a consistent, transparent and safe encrypting-decrypting behavior. It is ready for deployment in the nine ASD expert assessment centers in the Ile de France district.

Keywords. Autism Spectrum Disorder (ASD), Security policy, infrastructure policy, Cryptography, Access control, information system, privacy protection.

1. Introduction

Electronic health record and patient personal information protection are critical for the trust and reliability of an information system whether it is used within a locally accessible health care facility or over the Internet network. In that perspective, security policy and infrastructure were developed for the information system TEDIS [1, 2] dedicated to patients with Autism Spectrum Disorder (ASD) [3]. ASD patients' assessment and diagnoses in nine multidisciplinary centers in Ile-de-France district regularly produce a large amount of valuable clinical data. At the end of the assessment, ASD diagnoses are established. Schooling and health care measures recommendations in ambulatory units, specialized institutions and social care entities, are prescribed. TEDIS was designed to

[1] Corresponding author: Dr. M. BEN SAID, Service de Biostatistiques et d'Informatique Médicale, Hôpital Necker Enfants Malades, 149 rue de Sèvres, 75015 Paris, France; Email : mohamed.ben-said@ext.parisdescartes.fr

be used in daily care practice and for research. TEDIS manages systematic, exhaustive and continuous multi-centric ASD patients' data collection via secured Internet connections. It aimed at supporting medical community and public health decision-makers to improve knowledge about the psychopathological, physio-pathological and etiological processes involved in ASD, to follow-up incident patient cohorts and to support decision making, patient care and research.

The French national ethics and computing authority called Commission Nationale Informatique et Libertés (CNIL) authorized the use of TEDIS within a framework based on patient' consents [2], respect of privacy and protection of personal information [2]. The main philosophy was to limit, without affecting research goals, the risk of corrupting data integrity, of occurrence of unauthorized access and of establishing a link between patient identity and medical information. French CNIL formally requested to restrict access to authorized professionals and to store on the database server in a strongly encrypted form, information which directly or indirectly, leads to identifying patients e.g.:patient' last name, first name, date of birth and sex. On the other hand, TEDIS' main users among psychiatrics and clinicians, requested the ability to visualize patient' identifying information such as last name, first name, sex and the date of birth in the form of plain text to directly disambiguate patient' identity, facilitate linking clinical data to the appropriate patient, and facilitate data entry and quality control.

We obtained the French CNIL agreement in mid-2015. Bolstered by the CNIL authorization, medical experts, clinicians and interns in the child psychiatry department at Necker hospital started testing TEDIS application, over a private intranet network managed by the SBIM at Necker Hospital. The experimentation helped adapting and improving the system performance and stabilizing the conceptual data model. Information security management became crucial in the perspective of vulnerability of Internet access and we focused on strengthening our security approach. We will address in the following the security policy and infrastructure developed in TEDIS, to protect ASD' patients clinical data and patient' privacy. We will discuss the strengths and the limits of our approach and the future development of the project.

2. Literature review

Several articles address electronic health record security and privacy protection in the context of electronic health care information systems. Two major axes, with related subtopics of interest to our study may be distinguished: Security policy and Security infrastructure. "*The security policy should clearly define the guideline for creating, accessing and maintaining the integrity of patient e-health data and the scope of accountability for each responsible party. A security infrastructure covers the issues of login authentication, cryptography, access control,.., audit,.., and disaster recover"* [4]. ISO 22600-1:2014 *defines principles and specifies services needed for managing privileges and access control for communication and use of health information distributed across policy domain boundaries. It highlights information sharing between diverse health care entities … without compromising privacy and integrity of EHR* [5]. In former projects we accumulated experience in using ACL Discretionary Access Control (DAC) "*based on the decision of the owner, to limit access to patient data based on the identity of subjects and/or groups to which they belong*" [4]. We also were interested in Cryptography " cryptography goals *include privacy or confidentiality, data integrity, authentication and non-repudiation"* [4, 6].

3. Methods

3.1. Security policy

Hosting: A major part of the security policy relies on the hosting environment of the application and the database servers. French legislation defined personal health data hosting criteria in the decree n°2006-6 of Jan 4th 2006 [7]. We recently had a formal agreement with the informatics department of Assistance Publique des Hôpitaux de Paris (APHP), as an Accredited Hosting Provider in Health Personal Data, certified by the Public Health Ministry, to host TEDIS application and database servers.

Governance with regard to patient data access and use: A steering committee of psychiatric referents in each ASD expert assessment center, regional representative of health ministry, methodologists and physician proposed an ethical chart-agreement describing the commitment of the participants to the TEDIS project. The chart reminds the patient ownership of all patient data produced. It defines three contexts for users to access and use TEDIS' patient's data:

- Data production, within an ASD' assessment center where explicit nominative patient data are needed for quality control.
- Data analysis, for observing and following data production and quality control, within one center or between multiple centers. Only de-identified patient data are needed in this context.
- Research projects and exchange of data between projects, where specific scientific and ethics authorizations are required. De-identified patient data are used in this context.

The chart also reminds the roles of the scientific committee in selecting research projects, accrediting scientific publications, and the role of the steering committee in conducting and representing TEDIS' project.

Access to patient nominative data in the data production context: Authorized physicians and clinicians are designated by the psychiatric referent in each ASD assessment center to participate in TEDIS' ASD patients' data production. Each professional, may access only to the patient information s/he is responsible for, within one center. S/He has to adhere and sign the chart agreement and is responsible for the quality of the information produced in TEDIS. Centralized unique user name and password are issued and archived by the physician responsible of TEDIS at SBIM Necker. Formal individual paper correspondence communicates these keys to TEDIS' authorized users, along with the chart-agreement for signature and adhesion. A clinical research assistant supports TEDIS users, in the nine ASD centers, in data production and quality control.

3.2. Security infrastructure

Discretionary Access Control is implemented in TEDIS [4]. For acceptability and usability reasons, end users connect to the application using user name and password on TEDIS' welcome web page [8]. Additional transparent authentication controls are

performed. Network monitoring and number of connection failure limitation are implemented. They permit to detect abnormal behavior and robot phishing [9, 10].

Personal data protection: Scenario of use:

Before submitting the web page form, patient' nominative data are encrypted and sent over secured Internet connection to the web server. We used a strong symmetric encoding algorithm (AES with 256 bits keys) in JavaScript, which integrates the time of user' interaction with the web-page to generate unique key and cipher text [11]. We first tested storing the generated key and cipher text as such on the server databases. A security leak quickly appeared as we were using the same information on the server side and the web interface, e.g. a link between encrypted information and plain text identifying information could be made at the web page source code. We tried to hide encrypted information once the decryption occurred properly from the web page source code. The results were unstable and varied as function of the web browsers and the JavaScript versions.

We decided to strengthen the security at the server side, by encrypting again the data from the interface before storing it on the database server. To do this, we first encapsulated each cipher text and key into a specific Java Object using SealedObject Class [12] which enables creating an object and protecting its confidentiality with a cryptographic algorithm. We then serialized [13] the sealed objects and respectively saved them on the MySQL database server. To make it even more difficult for a non-authorized user to retrace relations between objects, we decided to use a patient-ID hash coding to map the objects in the databases [14].

Patient data retrieval: TEDIS' user selects a patient ID on the HTML web-page to request patient nominative and clinical information. Nominative information retrieval goes through the inverse process of encryption, object creation and storage. A first step would be to rebuild the patient ID hash code, then to select, de-serialize and unseal objects to be sent back to the user interface for decryption with the JavaScript program [11]. Patient' clinical information is directly selected from the database to the dynamic HTML web page. TEDIS' users may also consult a list of all patients within one ASD center. It implies that plain text nominative data of all patients in one center have to be restored and displayed properly along with clinical summaries. In this case, form fields' naming, for each patient record, have to be adapted for proper decryption with the JavaScript Program.

4. Results

Testing TEDIS in the local intranet environment followed the preconized security policy. It showed correct, consistent and transparent functioning of the ASD patient nominative data encrypting process for creating a new patient, retrieval or updating. 359 patient identities records were encrypted this way. Lists of patients within one ASD center display correctly, in transparent way decrypted identities and corresponding clinical summaries (the elapsed time from the web interface request to the display of the last patient record in a list is of 2266 milliseconds). Patient phonetic search on the web page identities list is also functional.

5. Discussion-Conclusion

Encrypting nominative data limits database patient identities search based on phonetics. Users either have to know the patient sequential identifying number or to select a patient record from a list of decrypted identities.

We are aware of the resources consuming of our solution, due to the multiple encryption-decryption and databases connections, storage and retrieval processes. We put most of the workload on the server side. This is particularly perceptible when requesting a list of patients with clinical summaries. The web-page JavaScript decryption process continues running sequentially while clinical information for all patients is ready for display. We acted on the database indexes and optimized queries to improve the server performance to obtain reasonable display latency.

The client desktop computer also needs to be performant to favor fluent encryption-decryption at the client side. We recommend using the latest version of web browsers supporting HTM5 (we use Firefox or Safari).

The privacy protection solution presented in this article showed consistency and stability compared to different, less elaborated solutions, tested in a first time, where we either had corrupted data on the server-side or a leak at the interface decryption.

The security policy, together with the security infrastructure provides a robust and reliable framework for an information system managing sensitive data, to be used in clinical setting and accessed via secured Internet connections.

References

[1] Ben Saïd M, Robel L, Vion E, Golse B, Jais JP, Landais P. TEDIS: an information system dedicated to patients with pervasive developmental disorders. Stud Health Technol Inform. 2010;160(Pt 1):198-202.
[2] Ben Said M, Robel L, Messian C, Craus Y, Jais JP, Golse B, Landais P. Patient information, consents and privacy protection scheme for an information system dedicated to pervasive developmental disorders. Stud Health Technol Informa. 2014; 205:755-9.
[3] http://www.dsm5.org last consulted on Oct 21st 2016
[4] Al-Hamdani WA, Cryptography Based Access Control in Healthcare Web Systems, in proceedings of InfoSecCD'10 Information Security Curriculum Development Conference pages 66-79, ACM New York, NY, USA, 2010.
[5] Jayabalan M, O'Daniel T, Access control and privilege management in electronic health record: a systematic literature review, J. Med Syst (2016) 40:261
[6] Tyagi N, Ganpati A, Comparative Analysis of Symmetric Key Encryption Algorithms, International Journal of Advanced Research in Computer Science and Software Engineering, Volume 4, Issue 8, August 2014, 348-354.
[7] Décret n° 2006-6 du 4 janvier 2006 relatif à l'hébergement de données de santé à caractère personnel et modifiant le code de la santé publique (dispositions réglementaires). JORF n°4, 5 Jan 2006, 174, texte n°14
[8] Czeskis A, Dietz M, Kohno T, WallachD, Balfanz D. Strengthening User Autentication through Opportunistic Cryptographic Identity Assertions. In proceedings ACM – conference on Computer and Communication security - CCS'12, pages 404-414, 2012, Raleigh, North Carolina, USA.
[9] Alzahrani BA, Reed MJ, Vassilakis VG, Resistance Against Brute-Force Attacks on Stateless Forwarding in Information Centric Networking, in ANCS '15: Proceedings of the Eleventh ACM/IEEE Symposium on Architectures for networking and communications systems, pages 193-194, 2015.
[10] Chenlang Lu, Zongda Wu, Mingyong Liu, Wei Chen, Junfang Guo. A Patient Privacy Protection Scheme for Medical Information Systems. J Med Syst (2013) 37:9982, pp 1-10.
[11] https://www.fourmilab.ch last consulted on Oct 21st 2016
[12] https://docs.oracle.com/javase/7/docs/api/javax/crypto/SealedObject.html last consulted Nov 6th, 2016.
[13] https://docs.oracle.com/javase/tutorial/jndi/objects/serial.html last consulted Nov 6th, 2016.
[14] https://docs.oracle.com/javase/7/docs/api/java/lang/Object.html last consulted Nov 6th, 2016.

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-333

Use of a Nationwide Personally Controlled Electronic Health Record by Healthcare Professionals and Patients: A Case Study with the French DMP

Brigitte SEROUSSI [a,b,c,1] and Jacques BOUAUD [b,d]
[a] *Sorbonne Universités, UPMC Univ Paris 06, INSERM, Sorbonne Paris Cité, Université Paris 13, LIMICS, UMR_S 1142, Paris, France*
[b] *AP-HP, Hôpital Tenon, Département de Santé Publique, Paris, France*
[c] *APREC, Paris, France*
[d] *AP-HP, DRCD, Paris, France*

Abstract. If the wide adoption of electronic health records (EHRs) is necessary to address health information sharing and care coordination issues, it is not sufficient. In order to address health information sharing, some countries, among which, France, have implemented a centralized framework with "new" nationwide care records. The French DMP is a centralized, nationally shared, electronic medical record, created according to the opt-in model. More than five years after the launching of the DMP project, DMPs have been created for 1.5% of the target population, which demonstrates the poor adoption of the tool by healthcare professionals. Among the 583,997 existing DMPs in June 2016, 41% were empty, and 24% of non-empty DMPs were actually accessed. If these "active" DMPs were equally accessed by both healthcare professionals and patients, patients accessed DMP documents four times more than healthcare professionals.

Keywords. Personal Health Records, Electronic Health Records/utilization, Attitude of Health Personnel, Patient Access to Records, Patient Participation

1. Introduction

In all countries, the increasing burden of patient care challenges the quality of care and even the sustainability of healthcare systems [1]. This is due in particular to the aging population since older adults are at higher risk of developing multiple chronic diseases and related morbidities, and the management of patients with multimorbidity that requires input from multiple healthcare providers across many care settings is complex. Numerous studies have reported that quality improvement strategies focused on the coordination of care reduced hospital admissions among patients with chronic conditions and emergency department visits among older patients [2].

A solution to improve care coordination is that the various care providers involved in the management of a given patient share the same consistent and accurate picture of

[1] Corresponding author, Département de Santé Publique, Hôpital Tenon, 4 rue de la Chine, 75020 Paris, France; E-mail: brigitte.seroussi@aphp.fr.

the patient's health. If the wide adoption of electronic health records (EHRs) is necessary to address health information sharing and care coordination issues, it is not sufficient. EHR information often remains stored in local health information systems and is not available to care providers outside the institution that produced the information. Mainly two solutions have been proposed to address this difficulty. The distributed model of clinical information sharing, adopted by the US, promotes health information exchange (HIE) between different EHRs [3]. The challenge of this model is to achieve actual interoperability between multiple EHRs. Other countries have decided to implement a nationwide, centralized, framework of clinical information sharing with "new" care records stored in platforms specifically created to support information sharing. These care records are created on top of the EHRs care providers use in their daily practice. They vary according to the countries in terms of content (care summary or extended documentation of the patient condition), of patient enrollment (opt in versus opt out), and of patient rights to access the medical record (only for care providers or personally controlled care records authorizing patient access). One of the most advanced examples of such a centralized nationwide system is the electronic summary care record (SCR) implemented in the UK [4]. Another example is the Australian personally controlled electronic health records (PCEHR) [5].

Close to the PCEHR, France has developed a nationwide care record for all patients, called DMP or "Dossier Médical Personnel" for Personal Medical Record [6]. The aim is to provide a tool that supports care coordination and promotes patient rights to support patient empowerment. More than five years after the undisclosed launching of the DMP (no marketing campaign), statistics of DMP usage show a very low adoption of the tool from healthcare professionals, but a true interest from patients.

2. Material and Method

The DMP is a centralized, nationally shared, patient-centered, electronic medical record. It is optional and free for all recipients of the French national health insurance. Only healthcare professionals can create DMPs. However, they must first get the patient's informed consent (opt-in model). Both healthcare professionals and patients may access the DMP online. Healthcare professionals involved in the management of a given patient need to be authorized by the patient to access the information recorded in his/her DMP. However, unauthorized health practitioners may "break the glass" and access the DMP without the patient's consent in case of emergency.

DMPs are not expected to be exhaustive. They are supposed to contain only the relevant information a healthcare professional considers necessary for other healthcare professionals in order to provide efficient and secure care, be it scheduled or not. A list of recommended documents has been proposed and includes hospital discharge summaries, encounter reports, radiology reports, biology reports, current prescriptions of drugs and care, patient care summaries (elaborated by GPs). DMPs also include a personal area where patients can enter any information they want, either because they think it could be of interest for healthcare professionals or because they wanted to have the information recorded in their DMP for their own use.

Currently, the DMP is essentially document-based and complies with the HL7 CDA release 2 standard. Documents may be categorized either as open, hidden, or sensitive. Open documents are accessible to the DMP owner and to all authorized healthcare professionals. Hidden documents are only accessible to the healthcare

professional that authored the document and to the primary care physician; all other authorized practitioners don't even know the existence of these documents (the hiding of documents is hidden to healthcare professionals). Sensitive documents are temporarily not accessible to patients because the information they contained is considered as sensitive, and it is recommended that patients be accompanied by their doctor when receiving this type of information.

The results presented in this paper are based on the study of DMP statistics provided by ASIP Santé (http://esante.gouv.fr/asip-sante/), the National agency in charge of managing the DMP project. We analyzed the distribution of DMPs according to the age of patients, DMPs' contents, and both healthcare professionals and patients accesses to the different types of DMP documents.

3. Results

At the end of June 2016, more than five years after the launching of the project in December 2010, there were 583,997 DMPs, which represents a coverage of 1.5% of the target population, made of 38 million of insured people.

3.1. Characteristics of Patients owning a DMP

Figure 1 reports the distribution of DMPs by class of age. DMPs for patients between 45 and 65 year-old (35%), and for those older than 65 year-old (39%) represent each more than one third of all DMPs.

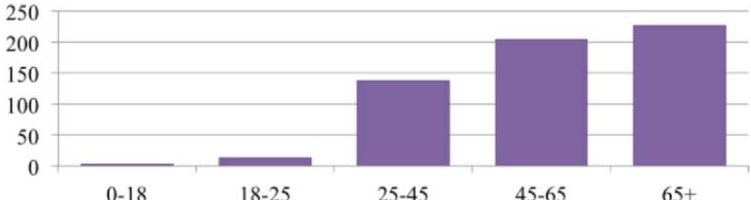


Figure 1. Distribution of DMPs (in thousands) by class of age.

3.2. DMPs' contents

DMPs contain a total of 2,390,123 documents. These documents populate 345,046 DMPs, with an average of 7 documents per DMP. Thus, 238,951 DMPs are empty representing 41% of the total. Figure 2 reports the distribution of the top 10 most frequent document types in DMPs. The most frequent is the report of medical consultation, which represents one third of all documents, the second is the hospital discharge summary (one fifth), the third one is the radiology report (one eighth). The other document types represent each less than 6%, while the top ten document types together constitute 93% of all contents. The patient-entered report, as displayed in Figure 2, represents 3% of all documents. There is a total of 137,954 patient-entered documents representing 5.8% of DMP documents. A total of 1,676 documents have been hidden to health professionals by patients (< 0.1%), and 43,759 sensible documents were masked to patients by health professionals, 26 times more than the documents hidden by patients, and 1.8% of all documents.

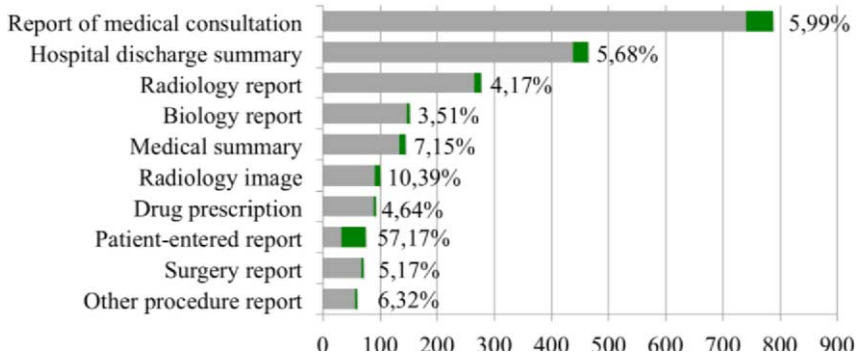


Figure 2. Distribution of the top 10 document types stored in DMPs (in thousands). Green parts correspond to accessed documents with their proportions figured.

3.3. Health Professionals and Patients Accesses to DMPs

Among the 583,997 DMPs, only 83,160 were accessed and considered as "active" DMPs, which corresponds to 14.3% of all DMPs and to 24.1% of non-empty DMPs. Among them, 97.9% were accessed by health professionals and 93,6% by patients. At the national level, 150,091 DMPs are "shared" meaning that more than one unique user accessed the DMP, either to push or to read documents. However, according to their types, accesses to DMP document may vary. Figure 2 shows for each document type a percentage tag providing the proportion of accessed documents per type of documents (figured in green). This illustrates that the patient-entered report is the most accessed document, more than half of all documents accessed (57.17%). Then follows the radiology image with 10.39% and the medical summary with 7.15%. Less than 6% of the reports of medical consultation, the most frequent document, have been accessed.

Health professionals and patients access to DMPs differently. Figure 3 provides the distribution of document accesses by category of users for the top 10 document types. On average, DMP documents were accessed by patients in 79.4% of the cases. This proportion reaches 98.1% for patient-entered documents. However, medical summaries, drug prescriptions, and other procedure reports were accessed more frequently by health professionals than by patients, 77.4%, 64.6%, and 54.3% respectively.

4. Discussion and Conclusion

DMP statistics show the poor adoption of the tool by healthcare professionals that are currently in charge of DMP creation. The opt-in model may have a role in these results as showed by the increase of PCEHR creation when Australia moved from op-in to opt-out[2]. If DMPs are created by healthcare professionals and mostly fed by them, statistics about DMP contents and accesses suggest a sustained interest of patients for their DMP. Patients produce their own documents, which represents a substantial 5.8% of DMP contents. As for DMP consultation, data shows that DMPs are more accessed by patients than by health professionals in proportions that go up to 80-20%. This

[2] https://myhealthrecord.gov.au/internet/mhr/publishing.nsf/Content/participation-trials

usage pattern suggests that with such a tool patients might be committed to their own care.

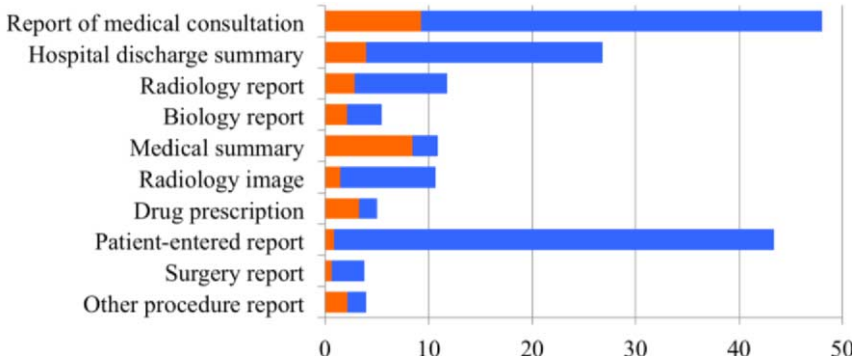


Figure 3. Distribution of the number of accesses to the top 10 document types stored in DMPs (in thousands) according to user categories. Orange parts correspond to health professional accesses and blue parts correspond to patient accesses.

A new French Health Act (01/26/2016) has promoted the relaunching of the DMP in order to drastically foster its adoption among health professionals and patients. Significant changes involve: (a) a semantic shift where the P of DMP means now "Partagé" or shared, instead of "Personal", to insist on its role as a coordination tool; (b) DMP governance by the French national health insurance (with financial incentives) instead of ASIP Santé; (c) patients' ability to create their DMP by themselves; and (d) automated push of heath insurance claims data into DMPs. The effect of such measures will have to be assessed in the coming years.

References

[1] Dall T, Gallo P, Chakrabarti R, West T, Semilla A, Storm M. An aging population and growing disease burden will require a large and specialized health care workforce by 2025. Health Aff (Millwood) 2013;32(11):2013–20).
[2] Tricco A, Antony J, Ivers N, Ashoor H, Khan P, Blondal E, et al. Effectiveness of quality improvement strategies for coordination of care to reduce use of health care services: a systematic review and meta-analysis. CMAJ 2014;186(15):E568–78).
[3] Rudin, RS Motala A, Goldzweig C, Shekelle P. Usage and effect of health information exchange: a systematic review. Ann Intern Med 2014;161(11):803–11.
[4] Cresswell K, Sheikh A. The nhs care record service (nhs crs): recommendations from the literature on successful implementation and adoption. Inform Prim Care 2009;17(3):153–60.
[5] Pearce C, Bainbridge M. A personally controlled electronic health record for Australia. J Am Med Inform Assoc 2014;21(4):707–13.
[6] B. Séroussi, J. Bouaud. Adoption of a Nationwide Shared Medical Record in France: Lessons Learnt after 5 Years of Deployment. AMIA Annu Symp Proc. 2016. To appear.

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-338

Use and Adaptation of Open Source Software for Capacity Building to Strengthen Health Research in Low- and Middle-Income Countries

Stefan HOCHWARTER[a,1], Salla ATKINS[a,b], Vinod K. DIWAN[a] and Nabil ZARY[c,d]
[a] *Department of Public Health Sciences, Global Health (IHCAR), Karolinska Institutet, Stockholm, Sweden*
[b] *School of Health Sciences, University of Tampere, Finland*
[c] *Department of Learning, Informatics, Management and Ethics, Karolinska Institutet, Stockholm, Sweden*
[d] *Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, Singapore*

Abstract. Health research capacity strengthening is of importance to reach health goals. The ARCADE projects' aim was to strengthen health research across Africa and Asia using innovative educational technologies. In the four years of the EU funded projects, challenges also of technical nature were identified. This article reports on a study conducted within the ARCADE projects. The study focused on addressing challenges of video conferencing in resource constrained settings and was conducted using action research. As a result, a plugin for the open source video conferencing system minisip was implemented and evaluated. The study showed that both the audio and video streams could be improved by the introduced plugin, which addressed one technical challenge.

Keywords. Capacity building, open source software, ICT4D, blended learning

1. Introduction

Efforts to strengthen health research capacity in low- and middle-income countries are needed[1]. The ARCADE HSSR and RSDH (African/Asian Research Capacity Development for Health Systems and Services Research/Social Determinants of Health) projects were two European Union-funded projects implemented from 2011 to 2015 and coordinated by the Division of Global Health at Karolinska Institutet in Stockholm. The projects aimed to strengthen health research across Africa and Asia by using innovational educational technologies[2]. The ARCADE projects can be divided into the following four interlinked components[3]. The first component of the ARCADE projects was the development and delivery of online courses on global health topics.

E-learning can be divided into synchronous and asynchronous education or e-learning. The latter one, asynchronous education, provides education from the teacher to

[1] Corresponding author, E-mail: stefan@hochwarter.org.

the student even if both are not online at the same time. Thus, in simple terms asynchronous education uses technologies and tools like e-mail, panels or e-learning platforms like Moodle. The strengths of asynchronous education lies within its flexibility on both time and place and its low requirements on the bandwidth, which is an important point for resource constrained settings[4]. On the other hand, asynchronous e-learning has also weaknesses, such as no possibility for immediate feedback, discussions on boards will last much longer on complex topics compared to face-to-face discussions and the lack of social interaction which may result in students not feeling connected to each other. Moreover, this method requires a severe discipline from the students as they need to manage their time of learning on their own[5]. To overcome the disadvantages of e-learning and to take into account bandwidth challenges, the ARCADE projects used mostly blended learning which is a combination of synchronous (face-to-face online or in person) and asynchronous (e-learning) methods.

The ARCADE project used different open source software to prepare, organise and deliver e-learning courses. Almost 55 courses were developed and delivered to over 920 postgraduate students in Africa, Asia and Europe using e-learning principles and specifically blended learning[6]. The open source e-learning platform Moodle was the main entry point for the students. Synchronous distance education was delivered with the help of minisip[2], which is an open source Session Initiation Protocol (SIP) implementation developed by ARCADE's project partner Royal Institute of Technology Stockholm[7]. However, as this partnership weakened over time, alternatives to minisip were evaluated and used. Content management was implemented with the help of Alfresco Community Edition, an Enterprise Content Management platform. Dissemination and publication of research findings were presented online at the self-hosted project's web site (www.arcade-project.org) which is powered by Wordpress.

Färnman et. al. investigated the challenges which the ARCADE project team came across during the four years of project runtime by interviewing 16 participants from 12 partner institutions. The main challenges for e-learning included problems in technology, availability of skilled technical staff across implementation sites and attracting students' interest in courses. The report also points out the high demand on bandwidth and software deficiencies in resource restrained settings[3]. This is specifically true for synchronous distance education. The limitation of the bandwidth, poor image and video quality as well as connectivity issues are challenges when video teleconferencing systems are used[8]. In this article we will demonstrate how open source software can be used to overcome bandwidth limitations.

2. Methods

The aim of this study was to evaluate how Open Source Software (OSS) can be used and optimised for distance education, particularly in the area of health research/education in a global setting. As the study took place within the ARCADE project which used the minisip software, the underlying study question can be stated as followed.

- How can OSS such as minisip improve the delivery of synchronous health education in resource-restricted environments?

[2] https://github.com/csd/minisip

The design of the study was defined by an action research framework[9]. Both qualitative and quantitative data were used, hence we followed a mixed research approach. This study can be divided in the different phases of action research.

1. **Observe:** A questionnaire was conducted with the aim to explore the ARCADE RSDH's teaching activities and its priorities to the system and quantitative data was collected at St. John's Research Institute in India.
2. **Reflect:** The data from the questionnaire and measurement were analyzed and possible improvements were identified.
3. **Act:** Based on the reflections, an improvement was chosen for implementation.
4. **Evaluate:** The implemented improvement was tested and evaluated at Wuhan's Tongji Medical College.

The questionnaire consisted of one multiple choice question and three free-text questions and was answered by ARCADE partners from Sweden (Karolinska Institutet) and India (St. John's Research Institute). A quantitative measurement of the Minisip performance using Wireshark was done at St. John's Research Institute in Bangalore.

The Minisip performance for the ARCADE RSDH project was evaluated and improvements were suggested, implemented and tested. The study took place at ARCADE partners in Sweden (Karolinska Institutet), India (St. John's Research Institute Bangalore) and China (Wuhan's Tongji Medical College). During the study, the focus of the terms "adaptability" and "performance" were narrowed down based on the outcome of the questionnaire.

Data were analysed using the statistical environment of R. The R package RQDA was used to organise and analyse qualitative data. Captured traffic from minisip was recorded and filtered by using Wireshark. Wireshark offers detailed analyse function for SIP calls and Real-Time Transport Protocol (RTP) streams, including package loss and jitter.

3. Results

In the first step (*Observe*), the background, aims and requirement of the underlying project was investigated by a questionnaire. A quantitative analysis of minisip at St. John's Research Institute in Bangalore showed that one challenge was the rather high Packet Delay Variation (PDV) or so-called "jitter". Based on the questionnaire in the first step and the measurement outcome (*Reflect*), we decided to implement a plugin for minisip with the aim to minimise the PDV. Minisip offers an interface to insert an extension at different stages of the traffic flow. The implemented extensions were inserted after the RTP pipeline at the project partner at Wuhan's Tongji Medical College, China (*Act*).

Two different algorithms were implemented to improve the traffic flow. As the audio packets all had the same size, a simple leaky bucket algorithm was sufficient. However, the video packets were of different size and therefore the byte-based token bucket algorithm was chosen[10].

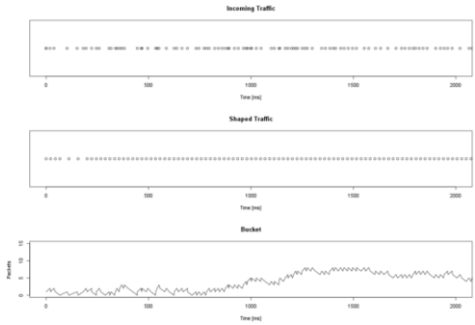


Figure 1. Measurement from the leaky bucket plugin for audio stream (Wuhan, China)

The final phase of the action research was to evaluate the outcome of the implementation. Figure 1 shows the output of the leaky bucket extension for incoming audio. The first diagram visualises the incoming packets, each dot represents one incoming packet (each 125 Bytes) before the extension is shaping the traffic. The second diagram shows the manipulated traffic, again one dot represents one packet. Finally, the third diagram represents the current content of the bucket with a maximum capacity of 15 packets.

Figure 2 illustrates the effect of the token bucket algorithm. The diagram "incoming traffic" visualises the incoming video packets over time and the diagram "shaped traffic" shows the resulted traffic after the token bucket extension. One dot represents one packet and the vertical axes shows the packet size. Two important parameters of the extension are also visualised: The packet queue shows the number of bytes (not packets) waiting for transmission and the last diagram represents the token available for traffic. One token is equal to one Byte in the used configuration.

As shown in Figure 1 and Figure 2, the PDV was successfully minimised for both audio and video streams.

4. Discussion

The study demonstrated with a specific example, how open source software can be adapted for capacity building in low and middle income countries based on a scientific background. However, to address the challenges which the ARCADE project team encountered during their four year project runtime, different strategies should be used. For example, investment in IT infrastructure and educating technical staff would be two possible strategies.

During this study we also faced a typical risk when using open source software that is not backed by a strong community. The main developer of minisip stopped his contribution and since then minisip is not further developed. The initial strength, namely the close partnership with the developer and thereby the possibility to receive an adapted software solution for our settings, turned out to be also a high risk factor.

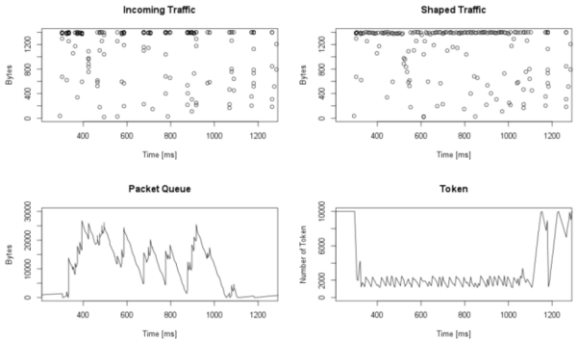


Figure 2. Measurements from the leaky bucket plugin for the video (H.264) stream (Wuhan, China)

Nevertheless, we believe that the use of open source software has high potential and strong advantages, especially in resource restrained settings. Within a limited time-frame, the project team successfully introduced an improvement for a specific problem based on measurements and interviews. This was made possible through an open source software that had a clean codebase and extensive documentation. Therefore external developers and contributors were enabled to introduce changes and improvements. As part of a future study, the plugins could also be integrated and tested in other SIP implementations. Finally, the implemented plugins could be further improved by automatically detecting the algorithms' parameters.

References

[1] McKee, M., Stuckler, D., & Basu, S. (2012). Where there is no health research: what can be done to fill the global gaps in health research? PLoS Med, 9(4), e1001209.

[2] Protsiv, M., Rosales-Klintz, S., Bwanga, F., Zwarenstein, M., & Atkins, S. (2016). *Blended learning across universities in a South–North–South collaboration: a case study.* Health Research Policy and Systems, 14(1), 67.

[3] Färnman, R., Diwan, V., Zwarenstein, M., & Atkins, S. (2016). Successes and challenges of north–south partnerships–key lessons from the African/Asian Regional Capacity Development projects Global Health Action, 9.

[4] Branon RF, Essex C. Synchronous and asynchronous communication tools in distance education. TechTrends. 2001;45(1):36-36

[5] Hrastisnski S. *Asynchronous and synchronous e-learning.* Educause quarterly. 2008;31(4):51-55

[6] Atkins, S., Marsden, S., Diwan, V., & Zwarenstein, M. (2016). North–south collaboration and capacity development in global health research in low-and middle-income countries–the ARCADE projects Global Health Action, 9.

[7] Rosenberg J, Schulzrinne H, Camarillo G, Johnston A, Peterson J, Sparks R, et al. *SIP: session initiation protocol.* RFC 3261, Internet Engineering Task Force; 2002. Available from: http://www.ietf.org/rfc/rfc3261.txt.

[8] Frehywot S, Vovides Y, Talib Z, Mikhail N, Ross H, Wohltjen H, et al. *E-learning in medical education in resource constrained low-and middle-income countries.* Human resources for health. 2013;11(1):1–15.

[9] Reason P, Bradbury H. The SAGE Handbook of Action Research: Participative Inquiry and Practice. vol. 2nd Reason P, Bradbury H, editors. SAGE; 2008.

[10] Tanenbaum A. *Computer Networks* 4th ed. Prentice Hall Professional Technical Reference; 2002.

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-343

Connecting the Links: Narratives, Simulations and Serious Games in Prehospital Training

Ilona HELDAL[a,1], Per BACKLUND[b], Mikael JOHANNESSON[b], Mikael LEBRAM[b] and Lars LUNDBERG[c]
[a] *Western Norway University of Applied Sciences, Faculty of Engineering and Business Administration, Norway*
[b] *University of Skövde, School of Informatics, Sweden*
[c] *University of Borås, Centre for Prehospital Research, Sweden*

Abstract. Due to rapid and substantial changes in the health sector, collaboration and supporting technologies get more into focus. Changes in education and training are also required. Simulations and serious games (SSG) are often advocated as promising technologies supporting training of many and in the same manner, or increasing the skills necessary to deal with new, dangerous, complex or unexpected situations. The aim of this paper is to illustrate and discuss resources needed for planning and performing collaborative contextual training scenarios. Based on a practical study involving prehospital nurses and different simulator technologies the often-recurring activity chains in prehospital training were trained. This paper exemplifies the benefit of using narratives and SSGs for contextual training contributing to higher user experiences. The benefits of using simulation technologies aligned by processes can be easier defined by narratives from practitioners. While processes help to define more efficient and effective training, narratives and SSGs are beneficial to design scenarios with clues for higher user experiences. By discussing illustrative examples, the paper contributes to better understanding of how to plan simulation-technology rich training scenarios.

Keywords. Prehospital training, simulation, serious games, user-experiences collaboration, narratives, design

1. Introduction

Today's healthcare cannot bear its cost, is not robust, does not satisfy a large number of patient needs and current development paths have to be changed [1]. Subsequently, education and training for the involved professionals need to be changed. These changes should provide a better understanding of the flows of knowledge, technologies, and financing in a larger context [2].

This paper illustrates planning and design of a large practical study providing contextual understanding and high user experiences for training prehospital work. Contextual understanding was planned based on narratives distilled with different, often collaborating professionals. Interviews and observations contributed to define

[1] Corresponding author: Ilona Heldal, Western Norway University of Applied Sciences, Faculty of Engineering and Business Administration, Norway, E-mail: ilona.heldal@hvl.no.

basic activity chains for their everyday practices. High user experiences were supported with using simulations chosen for the corresponding activities such as car simulators for driving, manikins or patient simulators for realistic patients' environment, and scenario simulators for simulating realistic environments for patient care [3]. To set up technologies and to provide engaging context through these activity chains, methods and solutions from serious games were used.

The aim of this paper is to illustrate resources needed, in terms of technologies, expertise and time for planning and performing simulation based training scenarios supporting contextual understanding and collaboration needed between different experts. The motivation behind this work is to contribute to more adequate preparedness, and to illustrate possibilities for collecting and providing evidences for contextual understanding needed for debriefing and assessments for healthcare professionals. The focus is on prehospital training and using simulations and serious games (SSGs). Some parts of this study can be found in earlier papers (see Section 3).

Such a qualitative study has several limitations. The exact amount of resources needed for realizing contextual training depends on several factors: e.g. the actual context, policies, capabilities, problem definitions, cognitive workloads, goals and technologies that always can be improved. Many of these factors varies from settings to settings and individual factors always can be improved. However, improving one factor does not necessarily mean improvement for the whole. Planned new training situations and designing new training environments have to deal with all factors influencing the whole chain from getting the alarm to driving the patients to the hospital and reporting the case. For prehospital work one may win important minutes in handling an alarm quickly and driving to the patients, but these minutes can be lost meeting difficulties in necessary activities when handling the patient or when driving him or her to the hospital.

2. Background

Organizations are providing their best to set up modern training and education environments. However, defining new settings for collaboration [4, 5], understanding interoperability [6], and finding technical acceptance for supporting technologies [7] is complex and requires an overall understanding of human-technology collaboration in the context, which this paper aims to contribute to. To develop better training methods for EMS personnel, as compared to previous methods, include training a safe and effective work flow since the whole activity chain is trained. Prehospital care is such a complex healthcare activity, where training traditionally has focused on particular details rather than the entire work flow.

Several papers argue for using lean [8, 9] and lean simulation [10] for healthcare education. Recognizing the benefits of lean especially for handling continuous flows and waste is important; however, implementing them in particular contexts can be difficult [11]. Examples from practice [9] or guidance for implementations [12, 13] are often missing. Lean is often associated with defining processes important to focus on strategies, progress, owners and patient goals [10, 14]. Lean and processes are not enough, though, to define necessary technologies [3], and scenarios [15] to begin with, or to know how to begin to implement SSGs [5]. Even though it is recognized that processes can support understanding of complex healthcare activities [16], in many process descriptions (e.g. [17]), the prehospital care is often treated as a single step in

the systematic care taking process. Sometimes the flow in training can be more important than training separately individual activities [3, 10, 18].
Earlier studies argue that considering narratives described by the practitioners can help [19]. Thus, defining a logical path guided by a process description, necessary and compulsory activities can be considered, then these activities can be enhanced and or supported by simulation technologies for naturalistic user experiences [5].

3. Towards Planning Contextual Studies by using SSGs

Emergency management stakeholders from one[2] of the twenty regions/county councils in Sweden, decided to promote the use of simulation technologies for training everyday activity chains[3] for professionals from the ambulance services. Motivated by the requirements from the practitioners the researchers behind this paper observed and recorded more than one hundred actual training situations where pairs or groups of three ambulance nurses trained in everyday settings. Other simulator centers were visited and different simulation based training possibilities with the use of SSGs were analyzed. A prehospital activity chain demonstration was used as background material for discussing possible technologies, scenarios, choices and consequences for training.

Based on lessons from the demonstration[4] and wishes from the practitioners a pilot study for training 24 nurses (12 pairs) was designed and performed. Each pair performed a simulation rich scenario with eight simulators and an everyday scenario, only with a patient simulator. The simulation rich scenario lasted approximately one hour and the everyday scenario 30 minutes, including the scenarios as such and follow-up activities such as interviews and questionnaires. The pilot study lasted 3 working days involving 23 researchers. The involved studies are described in Table 1.

Table 1. Overview of involved studies with main activities and main results.

	2012 Pre-Study	**2012-2013 Demonstration**	**2014-2015 Pilot Study**
Data obtained	Observations from real life (> 100 training situations). Interviews with instructors.	Observations from a single demonstration (2 nurses). Video recordings. Interviews with instructors and practitioners.	Observing 24 (12 pairs) nurses in a 3-day study by using a simulation rich setting compared with an everyday setting. Video recordings. Interviews with practitioners.
Main results	Lessons from actual training situations [20]. Understanding the gap between practical needs and technical possibilities.	Defining and showing how to set up technologies demonstrating activity flows needed to train [3]. Defining a technology supporting debriefing.	Defining a scoring instrument [21] and investigating the benefits of using immersive technologies and simulators [15].

[2] Västra Götaland Regional Council (VGR)

[3] These chains usually begin with a call from the dispatch center and driving to the patient. After arriving at the scene, initial patient assessment is performed, and if needed, the patient is transported by ambulance to the hospital. Ongoing treatment is made en route and finally, the patient is handed over at the hospital [20]. Ideally, the training ends with a debriefing.

[4] https://www.youtube.com/watch?v=WGCgEzNwXDU (Retrieved 2/22/2017)

According to our present knowledge, the whole activity chain with several realistic simulations is not trained in current education. For example, none of the simulator centers visited, or earlier research studies described prehospital training using patient simulators in an ambulance car simulator driving through a familiar city.

4. Connecting the Links: SSGs and Narratives to Train Activity Chains

The prehospital chain includes complex activities, not usually trained activities, e.g. finding/recognizing places from where the alarm was sent, handling the patients' environment, neighbors, pets, driving and care of the patients during the way to the hospital, using actual communication channels and technologies, discussing what way to drive, etc. A nurse taking care of the patient in the ambulance car sometimes has to practice five different duties in parallel [3]. Some of these activities are only discussed during classroom situations, and not trained at all.

To train whole activity chains requires many resources [9]. Planning new training needs coordination and collaboration from many different parties [2, 11], here: from different research areas, experts in technology development and use, instructors involved in educating healthcare professionals, responsible managers from the region, physicians and other healthcare professionals. Simulating rich settings, as live simulations is expensive. This can explain why whole activity chains are rarely trained. On desktop computers, however, SSGs promises benefits for contextual training or training decision making e.g. for incident commanders [5].

The demonstration with using simulators at the hand, convinced the practitioners to invest in the pilot study systematically examining the added value of a simulation-rich, high-end contextualized scenario compared with a traditional basic scenario. For this, all technologies were adjusted based on the process description for training and the high end simulation was enriched with the narratives described by the practitioners from the pre-study and demonstration. The practitioners appreciated the high-end settings, with the experiences gained. These contributed to a discussion of problems not previously trained in realistic settings, e.g. handling aggressive pets.

The pre-study and the demonstration also illustrated that practitioners need better evidences for debriefing. Therefore, a logging instrument was constructed, with tags connecting certain video images to the main activities. Certain predefined tags, such as the video sequence for entering a house was connected to checking securing activities in an instrument developed to support debriefing. By adding this feature, the instructor can, in the simulation-rich scenario, directly jump to a certain video during the debriefing to see if the ambulance nurses did necessary preparations for entering the patient's home. Such preparations are necessary for their own safety. Using this instrument, the nurses and the instructor(s) could follow the debriefing with recorded and ordered evidences grouped around main activities and time.

Representing all activities that the nurses perform on a daily basis was experienced to make the scenarios more trustworthy in contrast with traditional simulation training, where focus is on separate activities such as patient care or driving. The main benefit of process-based training approach was: choosing suitable simulators and necessary activities. In order to obtain increased experiences, the necessary activities can be completed with unexpected – but realistic situations. These situations are not necessarily known by designers, but can be defined from interviewing experienced practitioners. These situations and items, not necessarily connected to the patient's

medical condition, contribute to variation and makes the training for a medical condition different, less boring, and may also increase the willingness to train more regularly.

While this work exemplifies the benefit of varying situations, it also pinpoints the need for defining additional methodologies to understand the role of a basic, minimalistic scenario and how this can be supported or extended by technologies in systematic evaluations. A first step here is not only to provide seamless use of technologies, but also to support their seamless integration into education. Defining process description for training requires a good understanding of the context with focus on necessary actors, technologies, and activity flows. Practitioners do not necessarily know about the existence of many technologies, or how to choose and use them on an everyday basis.

References

[1] M.E. Porter and E.O. Teisberg, *Redefining Health Care: Creating Value-Based Competition on Results*. Boston: Harvard Business School Press, 2006.
[2] J. Frenk, et al., Health professionals for a new century: transforming education to strengthen health systems in an interdependent world. *The Lancet*, 2010, 376(9756), 1923-1958.
[3] P. Backlund, et al. Collaboration Patterns in Mixed Reality Environments for a New Emergency Training Center. *Proceedings of the IEEE European Modelling Symposium*. 2013. Manchester: IEEE.
[4] G. Fischer, G. and E. Giaccardi, Meta-design: A framework for the future of end-user development. *End user development*. 2006, Springer, 427-457.
[5] I. Heldal, Simulation and Serious Games in Emergency Management: Experiences from two case studies. Proc. of 22nd Int. Conf. of VSMM, 2016. Maleysia: IEEE.
[6] A. Sadagic, and F.A. Yates Jr, Large Scale Adoption of Training Simulations: Are We There Yet? *Proc. of The Interservice/Industry Training, Simulation and Education Conf.* (I/ITSEC), 2015, New Orleans.
[7] K. Suneson, K. and I. Heldal, Knowledge Barriers in Launching New Telecommunications for Public Safety, *Proc. of ICICKM,* 2010. Hong-Kong, 429-439.
[8] J.J. Waring, and S. Bishop, Lean healthcare: rhetoric, ritual and resistance. *Social science & medicine,* 2010. 71(7), 1332-1340.
[9] B. Poksinska, The current state of Lean implementation in health care: literature review. *Quality Management in Healthcare*, 2010. 19(4), 319-329.
[10] S. Robinson, et al., SimLean: Utilising simulation in the implementation of lean in healthcare. *European Journal of Operational Research*, 2012. 219(1), 188-197.
[11] R.M. Åhlfeldt, et al., Supporting Active Patient and Health Care Collaboration: A Prototype for Future Health Care Information Systems. *Health Informatics J*, 2015. 22(4), 839-853.
[12] J. Bhamu, and K. Singh Sangwan, Lean manufacturing: literature review and research issues. *International Journal of Operations & Production Management*, 2014. 34(7), 876-940.
[13] L.B.M. Costa, and M. Godinho Filho, Lean healthcare: review, classification and analysis of literature. *Production Planning & Control*, 2016:1-14.
[14] I. Heldal, L. Lundberg, and M. Hagiwara. Technologies Supporting Longitudinal Collaboration Along Patients' Pathway: Planning Training for Prehospital Care. *Proc. of ICIKM, 2015*, Bangkok.
[15] H. Engström, et al., The impact of contextualization on immersion in healthcare simulation. *Advances in Simulation*, 2016, 1(1).
[16] S.F. Mertens, Gailly, and G. Poels, Supporting and assisting the execution of flexible healthcare processes, *Proc. of the 9th Int. Conf. on Pervasive Comp. Technologies for Healthcare*. 2015, Istanbul.
[17] A. Lanz, M. Reichert, and B. Weber, Process time patterns. *Information Systems,* 2016. 57(C) 38-68.
[18] I. Heldal, I., and L. Lundberg, Simulation Technologies Supporting Collaborative Training for Emergency Medical Services Personnel, *Proc. of KES Int.* 2017, Algarve.
[19] A.D, Brown, P. Stacey, and J. Nandhakumar, Making sense of sensemaking narratives. *Human relations*, 2008. 61(8), 1035-1062.
[20] P. Backlund, Ambulansträningscenter [The Ambulance training center]. 2013, *Report from School of Informatics: University of Skövde*. Skövde, 80p.
[21] M.A. Hagiwara, et al., Measuring participants' immersion in healthcare simulation: the development of an instrument. *Advances in Simulation*, 2016, p. 1-9.

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-348

Designing an E-Learning Platform for Postoperative Arthroplasty Adverse Events

Ole Andreas KRUMSVIK, B.Sc[a,b,1] and Ankica BABIC, PhD[a,b]
[a] *Department of Information Science and Media Studies, University of Bergen, Norway*
[b] *Department of Biomedical Engineering, Linköping University, Sweden*

Abstract. This paper presents a mobile software application development for e-learning based on the adverse events data within the field of arthroplasty. The application aims at providing a learning platform for physicians, patients, and medical students. Design of user interface aims to meet requirements of several user groups concerned with the adverse events of the knee and hip implants. Besides the clinical patient data, the platform wants to include even electronic patient data as a result of self-monitoring. Two different modules were created, one for medical staff and one for patients, both divided into the knee and hip areas. Knowledge is represented in forms of statistics, treatment options, and detailed, actual adverse event reports. Patients are given a choice of recommendation for two main situations: '*about your diagnosis*', and '*what if you get a problem'* as advice and guidance during the postoperative rehabilitation. Expert evaluation resulted in acceptance of the concept and provided feedback ideas. The patient evaluation has also been positive. Implementation will mean that a high-fidelity prototype will be developed and tested in larger user groups (medical staff, patients).

Keywords. Adverse events, arthroplasty, HCI, e-learning, low-fidelity prototype evaluation, mobile application

1. Introduction

The occurrence of adverse events creates a lot of unwanted costs and distress for society. In Norway, from 2011 to 2015, 16838 compensation claims were filed. 5308 got compensated, which lead to 3,4 billion NOK in compensation costs. In addition to these costs comes patient suffering and absence from work, as well as diminished life quality. The leading field regarding adverse events in Norway was Orthopedics, with 37% complaints nationwide [1]. Currently, there are no known learning platforms or solutions that can help us prevent this problem concerning the field of Orthopedics. Therefore, we are suggesting a learning platform that can support several user groups (physicians, patients, and medical students) to learn from postoperative adverse events in Orthopedics, more specifically arthroplasty. The main motivation behind this learning platform is to utilize meaningful, useful data that can be used as educational material which is not being wasted. Efficient learning from and preventing adverse events can contribute to significant learning purposes. The idea is to base the learning platform on data that is generated through several medical sources, one of them being the application dedicated to safety reporting [4] and other relevant sources such as

[1] Corresponding Author: Ole Andreas Krumsvik, e-mail: oa.krumsvik@gmail.com

biomedical databases, patient electronic records, and the Norwegian Arthroplasty Register. By continuously using this data as educational material for the several above-mentioned user groups, the occurrence and severity of adverse events could decrease. Another beneficial effect is that the patient care can be improved and the excess costs reduced. An important factor would be to empower patients in recognizing an adverse event because the average number of hospital discharges is increasing, and therefore shortening hospital stays in most countries presumably leading to uninformed patients [2][8]. The typical length of stay after e.g. total hip arthroplasty is one or two days [6]. A similar learning platform for oncology has been developed in the U.S seeking to improve cancer care (CancerLinQ) which was done with great success by using data from 170 000 previously treated patients [3].

The prototype is divided into two modules, one for medical students and physicians and one separate module for patients. It is designed for mobile devices, which enables context-independent learning.

2. Method

The platform was created using multiple development methods. Firstly, a literature study combined with field studies have been conducted to define the information and user needs of the platform. Subsequently, a model of the architecture was created on the basis of this information, which is illustrated in the results section. Following that, a proposed design solution for the concept was created with the software *Axure RP Pro*. Lastly, the concept was evaluated through judgment of feasibility by two experts in biomedical engineering with emphasis on the module for medical students and physicians. The patient module was evaluated by a patient that has underwent hip replacement. All the users have participated in design [5] by suggesting their improvements or wishes for changes and adding some functionality. Additionally, to further assess the module, a mini focus group with 8 questions was conducted with a senior surgeon and four nurses.

3. Results and Discussion

A low-fidelity prototype of the e-learning platform for the physicians and medical students was made to assess the concept and aid further development. Additionally, a model of the platform architecture (Figure 1) was created in collaboration with other developers [4] and the medical staff at the Haukeland University Hospital in Bergen. The architectural foundation is based on various databases such as patient electronic records, registry, and safety reporting data with the results of information retrieval, patient self-reporting data, and biomedical databases maintained by the Biomedical Laboratory. All of these data sources have valuable information that should contribute to the safety of patients and design of better IT services. HL7 standards will be used to exchange, integrate, share, and retrieve data [7]. Developing two different sets of applications, namely *Safety reporting application* and *Safety learning* would make a great use of the existing data and contribute new data to the less covered aspects of safety reporting in line with the occurrence and adjudication of adverse events. The learning part (Figure 1) of the architecture is quite unique as it attempts to make applications for several user groups interested in learning about the safety. Those are

physicians, medical students, biomedical engineers, and patients as a completely new group. Data regarding postoperative patient care as well as the education can be enabled using the IT technology.

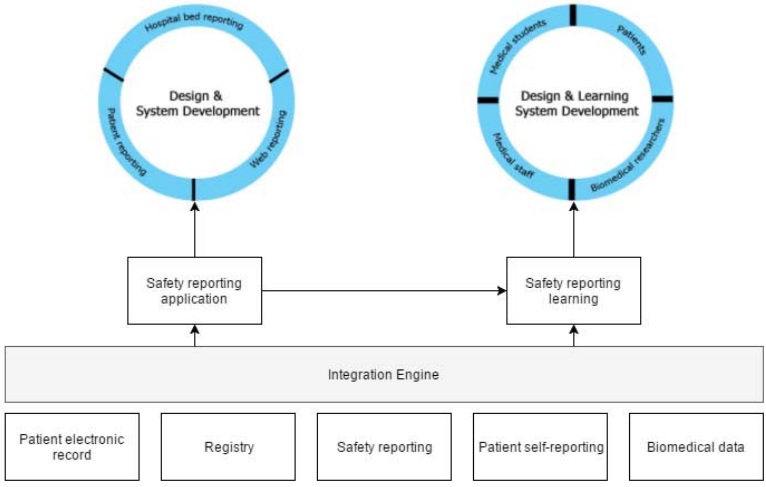


Figure 1: Platform architecture. Author's module – Design & Learning System Development.

Figure 2 (a) shows an overview of the knee. The red dots indicate an area of the knee in which a surgical procedure was performed. Figure 2 (b) presents a list of related adverse events in the given area that was chosen in Figure 2 (a). In this example sorted by severity. Figure 2 (c) displays three different options concerning the adverse event chosen in Figure 2 (b).

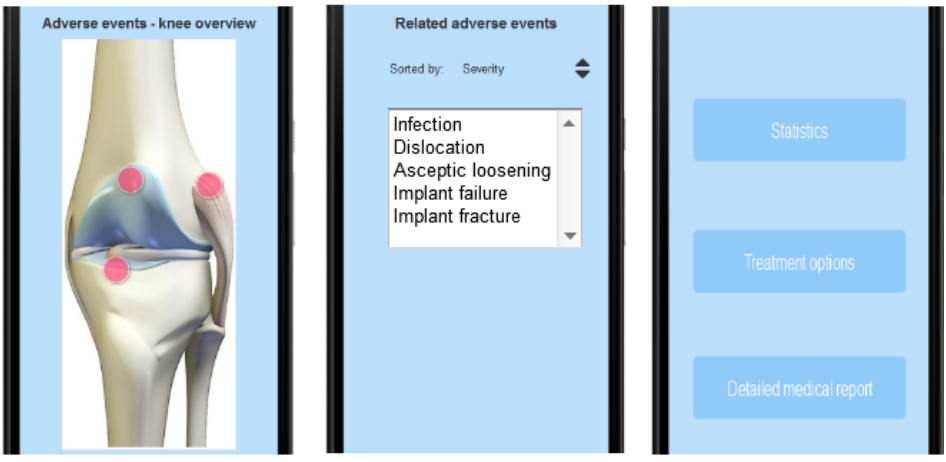


Figure 2. (a)Selecting an adverse event in the knee example (Screenshot); (b) List of adverse events sorted by severity example (Screenshot); (c) Three different options after selecting an adverse event from Figure 2 (b) (Screenshot).

The judgment of concept has been unreservedly positive. The clinician, a senior surgeon, has seen its potential to educate different types of patients and was very positive towards learning and using data for both students and patients. A profiling of the patient population that would be enabled by the common platform could be utilized to address patients' specific needs. One of the experts in biomedical engineering had considered the platform to be a very good concept, and definitely a feasible one. He had viewed it as a great use of existing new technologies, such as big data and internet of things. Furthermore, he suggested that it can be viewed as a tool for decision support making, and not solely a learning platform. He had provided suggestions on the design structure for further development. In particular, the design should give a possibility to add a rotatable model of the knee to the Figure 2(a), and to enable selecting several areas if the patient had multiple injuries. Additionally, in Figure 2(c) it could be reasonable to combine statistics and treatment options since each treatment would have a statistic associated with it. "Detailed medical report" would be better presented as "Case studies", with a list of cases that match the patient, adverse event, and treatment options. Lastly, he suggested to add a screen for patient details with basic options to cover the major risk factors such as age, sex, height, weight.

Presently, the patient module is design to contain relevant pre- and post-operation information, as well as relevant rehabilitation information to aid the patient throughout the whole treatment and to enable self-monitoring. One patient, a female, 51 years of age, who had recently undergone the hip replacement had evaluated the concept. Her assessment of the patient module was that it was *'a great idea'*. She appreciated that the information was easily accessible. By making the clinical and patient information available in any context, pre- or post- operatively, would make it beneficial for patients. The process would be easier both physically and mentally for the patient, especially during rehabilitation. The person has stated that the idea of self-monitoring with the help of this platform would be *'a great relief'*. Moreover, the focus group which consisted of a senior surgeon and four nurses from Linköping University Hospital, were very positive towards the module and emphasized its importance concerning patient safety.

4. Conclusion

Adverse events are difficult to manage from many points of view. The urgency of care, severity of cases and pressure to document the course of events are challenging for both medical staff and patients. It seems to be difficult to learn and utilize the knowledge of past events in a cost-beneficial way. This study has explored possibilities of using mobile and web technologies to define a concept and platform that would integrate data resources and create tools useful for medical staff and patients. The resulting conceptual design has been met with appreciation and several suggestions as how to improve the design solutions. Future steps will include designing the study cases, their presentations for physicians and students through different learning options. Patients' educational needs, and their options for self-monitoring, will be subject of design studies, as well.

Acknowledgments

The authors would like to thank Paul Johan Høl, PhD, Peter Ellison, PhD, and Geir Hallan, MD, PhD for their valuable insights into the design and the field based comments.

References

[1] Bergensavisen. (2016). Flere søker erstatning etter sykehusopphold. Retrieved September 6, 2016, from http://www.ba.no/nyheter/medisin-og-helse/helse/flere-soker-erstatning-etter-sykehusopphold/s/5-8-370537

[2] Eurostat. (2015). Hospital discharges and length of stay statistics. Retrieved September 12, 2016, from http://ec.europa.eu/eurostat/statistics-explained/index.php/Hospital_discharges_and_length_of_stay_statistics

[3] Miller, R. S. (2016). CancerLinQ Update. *Journal of Oncology Practice*, *12*(10), 835–837. http://doi.org/10.1200/JOP.2016.014530

[4] Åserød, H., Babic A. (2016). Designing a bed side application for adverse event reporting, Informatics for Health, April 2017, Submitted.

[5] Kensing, F., & Blomberg, J. (1998). Participatory Design: Issues and Concerns. *CSCW: The Journal of Collaborative Computing*, *7*(3–4), 167–185. http://doi.org/10.1023/A:1008689307411

[6] Erens, G. A., Thornhill, T. S., & Katz, J. N. (2016). Total Hip Arthroplasty. Retrieved December 7, 2016, from https://www.uptodate.com/contents/total-hip-arthroplasty

[7] Health Level Seven International (2016). Introduction to HL7 Standards. Retrieved December 7, 2016, from http://www.hl7.org/implement/standards/index.cfm

[8] Johansson, K., Salanterä, S., & Katajisto, J. (2007). Empowering orthopaedic patients through preadmission education: Results from a clinical study. Patient Education and Counseling, 66(1), 84–91. http://doi.org/10.1016/j.pec.2006.10.011

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-353

Understanding the Context of Learning in an Online Social Network for Health Professionals' Informal Learning

Xin LI [a,1], Kathleen GRAY [a], Karin VERSPOOR [a], and Stephen BARNETT [b]
[a] *Health and Biomedical Informatics Research Centre, University of Melbourne*
[b] *General Practice Academic Unit, Graduate School of Medicine, University of Wollongong*

Abstract. Online social networks (OSN) enable health professionals to learn informally, for example by sharing medical knowledge, or discussing practice management challenges and clinical issues. Understanding the learning context in OSN is necessary to get a complete picture of the learning process, in order to better support this type of learning. This study proposes critical contextual factors for understanding the learning context in OSN for health professionals, and demonstrates how these contextual factors can be used to analyse the learning context in a designated online learning environment for health professionals.

Keywords. Context analysis, health professional education, online social networks

1. Introduction

As medical knowledge expands and health care delivery becomes more complex, health professionals must commit to continuous learning to maintain up-to-date knowledge and skills. One approach to meeting their learning and development needs is through engagement in online social networks (OSN) [1]. OSN have been found useful to reduce professional isolation and support anytime-anywhere peer-to-peer interaction at scale. Also, they are thought to contribute to improving continuing professional development (CPD).

There are many OSN targeted towards health professionals but they appear to fail to support broader learning objectives. It has been recognised that there is a lack of understanding about how health professionals learn in OSN, making it difficult to design and manage this type of learning [2]. Understanding and evaluating the process of learning in OSN is important to realise the full potential of OSN for health professionals' learning [3].

Previous studies investigated learning behaviours by identifying the patterns of the interaction among health professionals [4, 5], and analysed textual dialogue among health professionals to understand how those dialogues support the process of learning [6]. However, there is insufficient attention paid to the understanding of learning context that is necessary to get a complete picture of the learning process in OSN [7].

[1] Corresponding Author: Xin Li, Health and Biomedical Informatics Centre, University of Melbourne, Parkville VIC 3010 Australia; E-mail: xinli87@gmail.com.

All learning occurs within a context. The nature of the context and how this context relates to the concepts being learnt has been widely shown to have an effect on learning outcomes [9]. Dey [8] gave the most widely accepted definition for context, which is "any information that can be used to characterize the situation of an entity" (e.g., a learner). The nature of learning occurring in OSN is self-directed and this requires health professionals to have self-regulated learning skills [10], but the ability to self-regulate learning is shaped by both personal-psychological and contextual factors [11]. Further, it is essential that these processes be learned 'in context' or in relation to the specific tasks of interest, in order for self-directed learning to be relevant to medical education [12].

This study aims to provide an understanding of learning context in OSN for health professionals' learning in two ways. First, through the review of literature, we identify the contextual factors that are critical for analysing the context of learning in OSN for health professionals. Then we use these factors to analyse context data collected from a designated online learning environment for health professionals. The intention of this research is to enable OSN operators to use such contextual factors to facilitate meaningful learning processes and improve learners' learning experiences in OSN.

2. Contextual Factors for Health Professionals' Informal Learning in OSN

The variety and volume of learning activity occurred in OSN demands a standard model for consistent analysis of context. Our literature review of context models previously proposed for online learning shows inconsistent approaches. Here, we identify the contextual factors (in Table 1) that we consider to be critical for analysing the informal learning context in OSN for health professionals, based on our synthesis of the context models in the literature [9, 13].

Table 1. Contextual factors for health professionals' informal learning in OSN

Context factor	Description
Learner factor	
Demographics	Age, gender, practice location, etc.
Work experience	Job role, year of experience, other professional activity, etc.
Education	Qualification, professional memberships, etc.
Learning interest	Specialised clinical interest, e.g. General Practice, mental health
Environment factor	
Time and Location	The learner's regular time and location of learning
Activity	The external learning activities that the learner is currently or will be involved in in future, apart from the interaction of OSN under study.
Relations	The social relation a learner has with other people outside of OSN, e.g. in their workplace.
Recognition	The ways of recognising learners' effective learning in OSN, e.g. through provision of CPD points
Application opportunities	Perception, and actual application of learned information or skills

3. Methods

3.1. Dataset

To understand the context of learning in OSN for health professionals, we conducted context analysis based on the contextual factors proposed above. The context data were

collected from the database of an online discussion forum provided by a health professional OSN host organization in Australia. The online forum was established in 2009 specifically for registered health practitioners and had more than 10,000 members. It was set up for doctors to discuss industry issues, share best practices and promote conversation within the health community.

For this study, we focused on the forum participants (N = 48) who contributed to forum discussions over three consecutive years from the period 2012 to 2014. Excluding two moderators, we collected the context data of 46 active participants. By conducting context analysis on the active participants, this study identifies common aspects of learning context that may be shared among active group in this community.

3.2. Procedures

We retrieved all data relevant to each contextual factor from the *user* table of the forum database via SQL Select statements. After performing data cleansing, we firstly analysed 'demographics' and 'education background' by using raw data (*date of birth, gender, practice location, university, college membership*), then we obtained context information for other contextual factors by processing raw data. We extracted and analysed the free text in data attribute *about me* to obtain context information for 'work experience' and 'learning interest. We also analysed the interaction history to obtain the preferred time and location the learners go online.

4. Results and Discussion

4.1. Learner Factors

Demographics. Most the participants were males aged 55+ years, which shows that mature health professionals are more actively participated in the OSN. This implies that it is important to consider the learning needs of older doctors when designing the learning in OSN. Using the Australian Standard Geographical Classification, we found that the majority of participants were practicing in a major city in the most populated Australian states (i.e. Queensland, New South Wales, and Victoria). These findings overturn assumptions that this type of learning is most relevant to professionals in regional and remote areas [14].

Work experience. 96% of the participants were General Practitioners (GPs). Of those, 56% were principal GPs and/or practice owners, implying that those with supervisory and/or management responsibility are more active in OSN. The distribution of years of experience is consistent with the age factor, which confirms that more experienced health professionals with 30+ years of work experience were more active in OSN for learning. Examining their practice status showed that 38% participants were retired or semi-retired, which reveal that even in retirement GPs, are willing to engage, and keen to using the OSN to maintain connection with peers and overcome isolation [15]. In addition, we identified 22% participants as having 'portfolio careers', that is, extra-professional activities. For example, they were also working as a medical officer at one or more hospitals, teaching or lecturing at a university, providing online consultations, involved in community services, etc. This shows that even very busy professionals were prepared to commit to learning in OSN.

Educational background. Most participants (72%) graduated in Australia, and 22% overseas. This suggests that the learning design of OSN should consider cultural and linguistic differences in the backgrounds of health professionals. Most participants (61%) are Fellows of the Royal Australian College of General Practitioners (RACGP), indicating that participants were more likely to hold an advanced qualification than be less qualified GPs.

Learning interest. All participants expressed major interest in general practice. In addition, more than half of the participants (54%) were interested in and/or had developed sub-specialties. The clinical areas of most interest among the participants were dermatology, women's health, chronic disease management, general medicine, diabetes, and sports medicine. This finding is somewhat consistent with a recent national GP survey that identified the top three clinical areas of interest to GPs as chronic pain management, cardiovascular health and diabetes [16]. This implies that GPs are keen to develop sub-specialised areas, and thus supporting their learning by facilitating more discussion on clinical topics is recommended.

4.2. Environment Factors

Although environmental context is critical to the design and management of learning in OSN [12], it was possible to collect only limited environmental data (i.e. time, location) from the dataset available. None of context data relating to learner activity, relations, recognition, and application opportunities were collected in this forum.

Time. By analysing the interaction history of the participants, we identified the time of day participants prefer to go online, revealing their self-directed learning schedules. We found that evening was more popular than morning or afternoon; some participants were online after midnight, or very early in the morning.

Location. Since evening was the preferred learning time for the participants, we can infer that a home office is likely to be their physical learning location. However, for those doctors who don't work during normal office hours (e.g. those who do shift work or work in hospitals) their physical location for learning online may not be in their home office but at workplace. Future implementation of the forum may consider tracking IP address of logged computers to obtain better understanding of their physical learning environment so appropriate learning content and activity can be suggested.

5. Conclusion and Future Work

It is important to understand contextual information about health professionals' informal learning in OSN, in order to better support this learning. This study suggested contextual factors that are critical for understanding learning context, and demonstrated how these contextual factors could be analysed, in the case of a small number of active participants in an online discussion forum. The findings are potentially useful to OSN operators aiming to increase participant engagement. The analysis of learning context help gain an understanding of these participants' experience, preferences and tasks, and thus contribute to developing more personalised and just-in-time learning for them.

The context data that were available in this forum were not complete, in particular environmental data. This study proved useful by identifying missing context information that would be worthwhile for the OSN operator to collect systematically in future, including learner activity, relations, recognition, and application opportunities.

This context information has important implications for the design and management of learning in OSN. In terms of the acquisition of this context information, not all could be collected automatically through the technical system; some information relied on manual input by participants (for example personal profile details). This could change in future if an OSN enabled participants to import relevant information about them from other online professional databases (for example, LinkedIn).

This study aims to provide an understanding of the learning context of health professionals in OSN. While the sample of our context data is too small to infer the general characteristics of the health practitioners in OSN, it shows a possible way for OSN operators to increase learner engagement by collecting and analysing the data of learning context in a designated online learning environment. In our future work, we plan to validate the findings of learning context, and use them to interpret the patterns of learning interaction and content in this OSN. Also, it is useful to understand how different contextual factors may influence the nature and outcome of learning in this OSN for health professionals.

References

[1] Cheston, C.C., T.E. Flickinger, and M.S. Chisolm, *Social media use in medical education: a systematic review.* Academic Medicine, 2013. **88**(6): p. 893-901.

[2] Institute of Medicine, *Redesigning Continuing Education in the Health Professions*. 2010, National Academies Press: Washington, DC.

[3] Sandars, J., P. Jaye, and K. Walsh, *Networked Learning in Continuing Medical Education: New Directions For The Evaluation of Effectiveness Perspective* in *Network-Based Continuing Medical Education: Social Media and Professional Development*, T. G., Editor. 2014, Nova Science: New York. p. 89-100.

[4] Stewart, S.A. and S.S.R. Abidi, *Using Social Network Analysis to Study the Knowledge Sharing Patterns of Health Professionals Using Web 2.0 Tools.* Biomedical Engineering Systems and Technologies, 2013. **273**: p. 335-352.

[5] Li, X., et al., *Analysing Health Professionals' Learning Interactions in Online Social Networks: A Social Network Analysis Approach*, in *Health Informatics New Zealand Conference*. 2015: Christchurch, New Zealand.

[6] Li, X., et al., *Discovery of Learning Topics in an Online Social Network for Health Professionals*, in *Health Data Analytics Conference*. 2016: Brisbane, Australia.

[7] De Laat, M. and B. Schreurs, *Visualizing Informal Professional Development Networks: Building a Case for Learning Analytics in the Workplace.* American Behavioral Scientist, 2013. **57**(10): p. 1421-1438.

[8] Dey, A.K., *Understanding and using context.* Personal and ubiquitous computing, 2001. **5**(1): p. 4-7.

[9] Tessmer, M. and R.C. Richey, *The role of context in learning and instructional design.* Educational technology research and development, 1997. **45**(2): p. 85-115.

[10] Sandars, J. and K. Walsh, *The Challenge of Creating Personalised Learning for Continuing Medical Education: A Network Perspective*, in *Network-Based Continuing Medical Education: Social Media and Professional Development*, T. G., Editor. 2014, Nova Science: New York. p. 21-33.

[11] Zimmerman, B.J. and D.H. Schunk, *Self-regulated learning and academic achievement: Theory, research, and practice*. 2012: Springer Science & Business Media.

[12] Cleary, T., et al., *Self-regulated learning in medical education*, in *Oxford textbook of medical education*. 2013. p. 465-477.

[13] Zimmermann, A., A. Lorenz, and R. Oppermann. *An operational definition of context*. in *International and Interdisciplinary Conference on Modeling and Using Context*. 2007. Springer.

[14] Brown, J., C. Ryan, and A. Harris, *How doctors view and use social media: a national survey.* Journal of medical Internet research, 2014. **16**(12): p. e267.

[15] Barnett, S., et al., *A virtual community of practice for general practice training: a preimplementation survey.* JMIR Medical Education, 2016. **2**(2).

[16] NPS MedicineWise, *National GP survey 2014*. 2014, NPS MedicineWise.

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-358

Ubiquitous Adoption of Innovative and Supportive Information and Communications Technology Across Health and Social Care Needs Education for Clinicians

Paula M Procter[1]
Reader in Informatics and Telematics, Department of Nursing and Midwifery, Sheffield Hallam University, UK.

Abstract. The paper presents the development, use and evaluation of an on-line undergraduate module delivering an academic-led programme of eHealth learning within nursing, midwifery, allied health professional and social work courses. The health information technology competency frameworks are explored along with an overview of the resulting module. The need for an academically led module will be made along with a description of the management required to maintain validity of content materials. A review of student evaluations will be presented. In conclusion the positive change in attitude and understanding of academic staff members towards health information technology through the inclusion of the module across all of the undergraduate courses will be explored.

Keywords. Education, health information exchange, informatics competencies, e-Learning, Nursing Informatics

1. Introduction

The Department of Nursing and Midwifery at Sheffield Hallam University is one of largest providers of nursing and midwifery education in England with some 700 nursing and midwifery students enrolling each year. The three year course for nurses and midwives is delivered 50% within the University and 50% within practice and the successful students complete the course with an academic award (BSc) and a professional award (Registered Nurse (RN)).

In information terms, nurses and midwives are the professional groups who have the most interaction with patients, their relatives and friends. This places nurses and midwives in a unique role amongst clinicians as the vital element link in the information infrastructure within health care. Since the early 1980's there have been many arguments forwarded that confirm the need to prepare clinicians to take an active

[1] Corresponding author, Paula M Procter, Department of Nursing and Midwifery, Faculty of Health and Wellbeing, Sheffield Hallam University, Sheffield S10 2NA, UK. E-mail: p.procter@shu.ac.uk

role in the development and use of information and communications technology (ICT), an example is that of Berg [1] in 1982 when she said:

> *"The choice is there and the time to make the choice is now. The decision must be whether to act traditionally and have change thrust upon the profession [nursing] from the outside or to anticipate this revolution in nursing practice, familiarize nurses with it, and prepare them to take an active part in the introduction of computers into the nursing community"*.

It is clear that a decision was not made across the wider nursing bodies at the time and it is only now that the importance of informatics engagement by clinicians has come to the fore through the publication of the Five Year Forward View [2], aligned documents [3, 4] and Leading Change, Adding Value [5] the framework from Jane Cummings, Chief Nurse, NHS England. It is acknowledged that there has been a plethora of suggestions for informatics competencies produced [6, 7] over the last ten to fifteen years but not one has been widely adopted to meet health professionals' needs.

In 2009 Systems of eCare was introduced as an academic-led fully on-line module using a managed learning environment within the Department of Nursing and Midwifery as part of the three year undergraduate nursing course with the following objectives:

- to understand, improve, influence and use new technologies and informatics, including remote care;
- to find the most reliable sources of information to support evidence based practice;
- to engage and guide patients through publicly available information sources;
- to incorporate ICT into patient consultations;
- to manage the nurse patient relationship when the nurse is not physically in the same place as the patient;
- to perform a quick and accurate data entry at the point of care;
- to understand the legal and ethical issues associated with managing and sharing patient information;
- to extract data to support decisions and monitor the outcomes of practice;
- to understand the role of technology in the delivery and organization of care
- to train other users such as patients and carers how to use relevant ICTs [8]

2. Systems of eCare

Systems of eCare 'docks' alongside the current curriculum at points determined by the student's stage of learning, in this way it does not require curriculum re-design. The module is made up of six units, each unit amounting to some 10 hours of study carried out on-line by the student at a time and place convenient to them. The intention being that a student completes one unit per semester (where there are two semesters per year). At the end of each unit, the student undertakes an on-line assessment, the result of which is contained in a printout which the student keeps in their portfolio. The outcome from the unit(s) is one of the issues discussed with the student during their regular Academic Advisor (personal tutor) sessions.

The design of the module uses an asset based model [9] which allows for ease of update is this ever changing domain. The six units cover issues such as health

information exchange, big data and population health, coding and classification, interoperability, information mapping, information intelligence, telematics, pharmaco-genomics, nano-technologies and robotics in health care. Each unit is routinely reviewed and updated every three months by knowledgeable health informatics experts to retain meaning and currency. Any significant policy change or innovative development is announced within 24 working hours for all cohorts.

Since the early beginnings in 2009, there are now over 2,500 nursing, midwifery, allied health professional and social work students with 24/7 access to the module. Figure 1 shows the monthly access hits between September 2015 and September 2016.

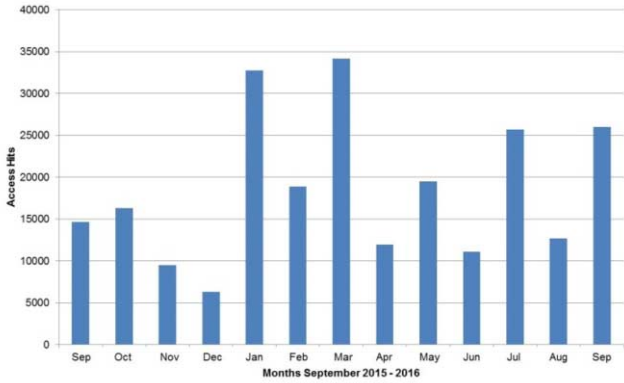


Figure 1. Access hits by students by month

Although access is available all the time, evidence has shown that whilst the students are on placements the number of accesses do reduce by some 43% per month. The module is optional and yet the rate of successful completion of all six units is considerably higher than might have been expected as shown in Figure 2. The dates are the cohort start dates for their three year undergraduate course.

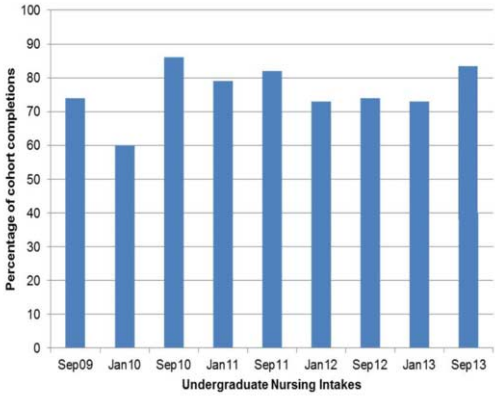


Figure 2. Student completions of all six units

2.1. Structured Evaluation - Key Findings

In September 2015, following an initial pilot to determine the validity of statements, a formal on-line evaluation survey using statements and Likert format response options commenced across the cohorts who had completed all six units. The following results are taken from the students anonymous evaluations for the September 2013 cohort (n=71).

As Systems of eCare sat outside the main curriculum a key question was asked to determine the students' view of where they saw the module in relation to their main course, 76% strongly agreed or agreed with the statement 'It is clear to me how Systems of eCare forms an integral part of my course' with 24% disagreeing or strongly disagreeing. 80% either strongly agreed or agreed that Systems of eCare was intellectually challenging with 63% stating that they found the content and resources in Systems of eCare supported their learning in other modules and assignments. When asked whether the student has been able to apply the knowledge they gained through completing Systems of eCare in their practice 74% strongly agreed or agreed with 23% disagreeing or strongly disagreeing and 1% unanswered.

In response to the statement 'I found that being able to access the online materials at a time that was convenient to me helpful' 96% strongly agreed or agreed with 3% disagreeing or strongly disagreeing and 1% stating 'not applicable'. This is supported by the metrics gathered as part of the managed learning environment as shown in Figure 3.

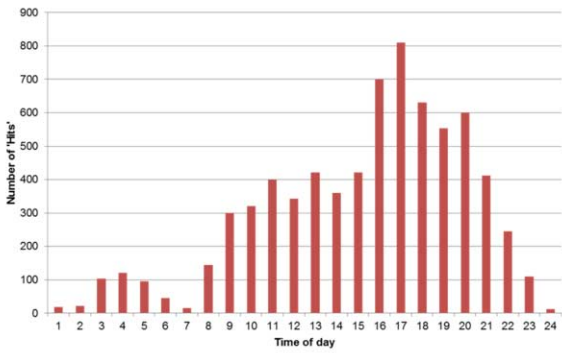


Figure 3. Hits by time of day, one cohort, one month

To measure the effectiveness of the design navigation and structure of the on-line module 56% responded that they did not need to contact the Systems of eCare Tutor at any time during their course.

2.2. Structured Evaluation - Student Comments

The comments ranged from '*Do we really need to know all of this?in such detail? Some of it is very relevant and important but I think 1 lecture on things such as data protection/freedom of information/ computer systems etc would be more effective but i imagine you have to put a certain amount of dedicated time to it, in which case the e learning is far better.'* to '*I found this to be a great learning chance to develop knowledge about the NHS'* and many points in between.

3. Discussion

It is often difficult to motivate students to engage with learning that is an optional extra to the course yet the results from the student evaluation survey suggest otherwise. The metrics in particular point towards significant access and use of the content. These two points will be considered in further depth during the presentation.

Each undergraduate student is allocated an Academic Advisor for the full period of their course. In the deployment of Systems of eCare it was vital to support academic colleagues in gaining greater understanding of the local and national policies for health and social care ICT. To foster this understanding each Academic Advisor receives regular Systems of eCare progress updates for their students along with an overview of the current unit's aim and learning objectives. With the increased confidence and support through Systems of eCare academic colleagues are now including a wider range of health informatics and telematics in their teaching.

4. Conclusion

Systems of eCare is not compulsory for the students and yet judging by the successful completions it appears to be meeting the students' needs, enabling them to have a robust grounding in health and social care ICT. Through this dynamic academic-led module nurses, midwives, allied health professionals and social workers have been, and continue to be, educationally prepared ready to be active participants in the development and appropriate use of innovative ICT through their professional careers. Future development and research plans will be highlighted.

References

[1] C Berg, The importance of nurses' input for the selection of computerized systems. In: Scholes M, Bryant Y, Barber B, editors. The impact of computers on nursing: An international review. Amsterdam, Netherlands: North-Holland; 1983. 42-58.
[2] NHS England, The Five Year Forward View. 2014
[3] NHS England, The General Practice Forward View. 2016
[4] NHS England, The Five Year Forward View for Mental Health. 2016
[5] NHS England, Leading Change, Adding Value, A framework for nursing, midwifery and care staff. 2016
[6] HIMSS, The Tiger Initiative. http://www.himss.org/professionaldevelopment/tiger-initiative. 2016
[7] HITCOMP, Health IT Competencies. http://hitcomp.siframework.org/ 2016
[8] P M Procter, Advancing Information and Communication Technology Knowledge for Undergraduate Nursing Students. *NI 2012: Proceedings of the 11th International Congress on Nursing Informatics*, (2012), IOS Press, 326-330.
[9] P M Procter, An Asset Based Model for Postgraduate Education. In: H.-A. Park et al (Eds), *Consumer-Centred Computer-Supported Care for Healthy People: Proceedings of NI2006*. IOS Press, ISBN: 1-58603-622-X, (2006), 167-171.

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-363

Mobile Medical Apps and mHealth Devices: A Framework to Build Medical Apps and mHealth Devices in an Ethical Manner to Promote Safer Use – A Literature Review

Mary SHARP[a1] and Declan O'SULLIVAN[a]
[a] *School of Computer Science and Statistics, Trinity College, Dublin 2*

Abstract. This paper presents a preliminary literature review in the area of ethics in the development of Mobile Medical Apps and mHealth. The review included both direct health apps and also apps marketed under the area of well-being in addition to mHealth devices. The following words and combinations of them were used to carry out the search for publications, mHealth, Apps, Ethics. The search engines used were Google Scholar, and PubMed. The paper is restricted to publications since 2012. The total number of papers found was 1,920 of which 84 were reviewed. The reason for so few being reviewed was that the majority only considered security. The search revealed many papers dealing with security for all types of apps and mHealth devices but there are very few papers dealing with the ethical issues related to Apps or mHealth devices in the area. It is noted however that the number of apps is increasing in number exponentially and therefore it is argued that it is necessary to pay attention to the ethical aspects. There are now estimated to be 165,000 apps available in this area. How ethics are addressed in health and well-being apps is important as they can have an effect on the health of the individual using them. In a similar way, the need for addressing ethical issues for development of well-being apps is evident. In a study [1] it was noted that even though Electronic Health Record (EHR) was the highest ranked tablet-related task only one third of clinicians said that EHR was optimized for smartphones. When apps are integrated with the EHR they fully optimize productivity. In the same study the significant challenges identified included the method of evaluation and selection of mobile health solutions in order to ensure that clinical outcomes, care and efficiency are included. Security is mentioned but again wider ethical issues were not a consideration. From the literature review it is clear that there is a need for guidelines for how developers of medical ad well-being apps and mHealth devices should address ethical issues during development, and the generation of these guidelines is the subject of ongoing research by the authors.

Keywords. mHealth, Apps, Ethics

[1] Corresponding author, School of Computer Science and Statistics, Trinity College, Dublin 2, Ireland, Email: mary.sharp@scss.tcd.ie

1. Introduction

According to the Oxford English Dictionary Ethics are moral principles that govern a person's behavior or the conducting of an activity. Ethical issues can arise in relation to the devices themselves, the software and the users. Examples for the devices include: do they work as intended, for software has it been thoroughly tested and for users what are the user interface design and privacy issues. Ethics in the area of IT in general is lacking and in the development of Apps is non-existent. In June 2016 the American Medical Association, at their Annual Meeting, [2] approved a set of guidelines for the ethical use of telemedicine. These were based on the fact that medial ethics should not be sacrificed for technology. There are a variety of smart phones and other devices available now all with significant power and memory. Fitbit is an example of another device which is designed for this monitoring eHealth. While the information passed between mobile phones, like wired telephones, is on a one to one basis this is not the case with many apps that store information about the users in the phone for further use. However there are not many rules, regulations or laws covering the use of, or even sale of, personal data collected by the App suppliers [3]. Prior to smartphones the most searched for items on the Internet were health related. This has now translated into a significant high rise of health apps being made available. Although a recent review [6] identified recommendations for mHealth applications in the areas of Privacy and Security, the comprehensive review did not include any references to other Ethical issues in the area. Apps have a set of "Terms and Conditions" attached but very few people actually read them before downloading. This opens up the users to numerous ethical dangers.

In section 2, the paper outlines the scope of areas that require Ethical consideration. Section 3 discusses the need for ethical guidelines for technology development in the area. Some of the ethical issues related to the usage of Apps are discussed in Section 4. Section 5 argues for the need for further development of ethical guidelines for apps and devices in the health and well being areas.

2. Ethical Scope

2.1. Health and Well-being Apps

The use of Apps in Health can be divided into three areas from an Ethical point of view, Apps with indirect health implications, which are copies of established text books, search engines for retrieving up to date publications or pharmaceutical catalogues. They are based on established facts that have been peer reviewed. These apps are mainly used for reference, training and education [4].

The second area is apps with direct health implications, apps for diagnosis, collection of health data, decision support, medical imaging and calculation of dosage for drugs. Associated with these there are many ethical issues for example accuracy of diagnosis using decision support or imaging, incorrect calculation of dosage. Healthcare is increasingly being affected by migration and the need to communicate in various languages. Smartphones can be used for translation between speech and text [7].

The third area is those apps used for patient monitoring. Examples of Apps available are from those for measuring clinical blood levels which are able to communicate with an EHR, to Apps that detect falls or lack of movement in the elderly.

This can raise ethical issues if incorrect information leads to incorrect prescribing. If a patient constantly drops a device and a false fall alarm is detected this may lead to the monitor not responding to a genuine fall later. The latest monitoring device to come onto the market is a biosensor. This is a patch that can monitor vital signs or certain external conditions like sunlight. It is in the form of a lightweight patch, likened to a patch to help smokers stop smoking. The first patch, Vital Patch, was unveiled at HIMSS16 in March 2016. This patch measures single lead ECG, heart rate, heart rate variability, respiratory rate, skin temperature, posture, step count and fall detection. The patches are disposable. They raise ethical issues in the area of collecting personal information and then being discarded with the information.

3. Ethical Guidelines for Design and Development

3.1. Requirement for Guidelines

Designers of Health and well-being Apps and medical devices must be aware of the consequences of errors in the development. This is not only an education issue but we argue one that requires guidelines for developers. Thus it is important to be clear on the requirements such a set of guidelines need to meet. As medical research is constantly bringing in new treatments, the systems have to be updated in line with medical progression. All potential uses of the system must be explored to ensure they cannot be used in a non-ethical manner. The guidelines for a developer will have to cover all ethical concerns in an easy to follow manner. They will include reference to the legal and non legal requirements. The primary concern will be for the safe development of the apps/devices so they will do no harm to the individual using them or to those who rely on the data generated. When building ethical guidelines into the mHealth and App Development lifecycle they cannot make the lifecycle more cumbersome to use.

3.2. Ethical Issues Associated with the Storage of Data

If the data is going to be stored in the cloud extra Data Protection Regulations will apply to ensure that ethical issues are avoided. In Europe the data must reside on a server within the European Union. Therefore it is necessary for those using the cloud to know the terms and conditions under which the cloud operates. If the data is to be combined with other data for research then the origin of the data needs to be known. Protection of data from being sold to a third party must also be guarded against. Some privacy policies are very loose or in some cases do not even exist. Many apps require a lot of personal details to be included which is then saved in a place that is unknown to the user. All of the information, including non medical data, has a market value [8].

4. Ethical issues related to the Use of Smartphones and Apps

4.1. Health Care Organisations

Health Care Organisations are using apps and mHealth devices increasingly for a variety of reasons. Reminders about appointments are in widespread use. One area of

ethical concern associated with this is to ensure there is follow up if patients do not respond. Ethical issues arise when a Health Care Organisation decide to implement a policy on the type of apps and mHealth devices to approve for the various functions they perform to support their overall work. All apps should be thoroughly tested by the organisation before being deployed and only those approved by the organisation should be employed. The best know app in the area of health support is Epocrates [9], this is now in version 16.6 and so it can be said that this has been tried and tested over time. Epocrates comes in two versions one a free one and also a subscription one each with a wide range of uses.

4.2. Primary Care

A review [10] revealed that 90% of physicians use smart phones in their professional life. One third of the physicians have recommended an app to patients [11]. The main reason for not recommending Apps is the lack of regulation by the FDA [12]. In 2013, the FDA came out with regulations and rulings of what apps are deemed appropriate as health apps and published its ruling (US Food and Drug Administration, 2014) [13]. The document states that "Only apps that serve as medical devices or transform a device into a medical device or perform patient - specific analysis and then provide a diagnosis or treatment on the basis of it will be regulated". The Health Informatics Unit of the Royal College of Physicians produced guidelines in April 2015 on using Apps in Clinical Practice [14]. The main theme of the guidelines are that even if the App has a CE mark it does not necessarily mean that it meets best practice. If an App is used for any medical purpose it is classified as a Medical Device. If an App uses patient specific information then it needs a CE mark. However ethical issues in the areas of accuracy, security and privacy are still issues for most health apps and the use of them will be limited by physicians.

4.3. Images

Prior to the introduction of smart phones all medical images were typically taken by professional clinical photographers. Clinical photographs can be used for a variety of purposes including diagnosis, treatment, education, research and medical legal situations all of which carry ethical issues with them, including identification of a patient, Data Protection, and ensuring that the image is clear to enable a valid diagnosis. It is important that informed consent is received from the patient for using the image. All of these aspects are requirements under the Data Protection Legislation [15].

4.4. Decision Support Systems

There are also ethical concerns particularly in relation to medical or wellness apps for Decision Support. One of these concerns is in the area of the accuracy of the app to do what it is meant to do. This can manifest itself in both apps used for monitoring and also for those used for Clinical Decision Support. A study [16] investigated the use of Apps by junior doctors for advice in the absence of a senior colleague being available. It found that the junior doctor will not necessarily know if the information supplied is accurate.

5. Conclusion

Having undertaken the preliminary literature review, it is clear that there is a need for guidelines for technology developers to consider and address ethical issues during. A Framework for incorporating ethics into the development of Health Apps and mHealth devices is being developed by the authors of this paper and will be based on the following definition, "a framework generally provides a skeletal abstraction of a solution to a number of problems that have similarities. A framework generally outlines the steps or phases that must be followed in the implementation of a solution without getting into details of what activities are done in each phase" [17].

Acknowledgements

This study is partially supported by the Science Foundation Ireland (Grant 13/RC/2106) as part of the ADAPT Centre for Digital Content Platform Research (http://www.adaptcentre.ie/) at Trinity College Dublin.

References

[1] Athenahealth (2014). *"Epocrates Mobile Trends Report*."
[2] https://www.ama-assn.org/ama-adopts-new-guidance-ethical-practice-telemedicine accessed 4th November 2016.
[3] Chahrour, N., *Ethics of Health Intervention Strategies on Musculoskeletal Disorders and Public Health*. Thesis University of Toronto
[4] Maged N. Boulos, K et al. *Mobile medical and health apps: state of the art, concerns, regulatory control and certification* Online Journal of Public Health Informatics 5(3): e229, 2014
[5] Digital Health Intelligence Limited, 9th June 2016
[6] Martínez-Pérez, B, De La Torre-Díez, I et al. *"Privacy and security in mobile health apps: a review and recommendations."* Journal of medical systems 39 (2015): 1-8.
[7] Baumgart, D.C., 2011. *Smartphones in clinical practice, medical education, and research*. Archives of internal medicine, 171(14), pp.1294-1296.
[8] Dembosky, A. (2013). "*Pregnancy apps raise fresh privacy concerns*." Financial Times.
[9] http://www.epocrates.com accessed 4th November 2016
[10] Boruff, J T., Storie. D "*Mobile devices in medicine: a survey of how medical students, residents, and faculty use smartphones and other mobile devices to find information*." Journal of the Medical Library Association: JMLA 102.1 (2014): 22.
[11] Hussain, M, et al. "*The landscape of research on smartphone medical apps: Coherent taxonomy, motivations, open challenges and recommendations*." Computer methods and programs in biomedicine 122.3 (2015): 393-408.
[12] Gauntlett, C., MacCarthy, J., et al 2013. *Patient apps for improved healthcare: from novelty to mainstream*. Parsippany (NJ): IMS Institute for Healthcare Informatics.
[13] Nathan G. Cortez, J.D., et al. Glenn Cohen, J.D., and Aaron S. Kesselheim, M.D., J.D., *M.P.H. FDA Regulation of Mobile Health technologies.* New England Journal of Medicine, July 2014, 372-379
[14] https://www.rcplondon.ac.uk/guidelines-policy/using-apps-clinical-practice-guidanceCommission, E. (2012). "*Guidance document - Qualification and Classification of stand alone software." Medical Devices.* Accessed 4th November 2016.
[15] Palacios-González, C *The ethics of clinical photography and social media Medicine, Health Care and Philosophy* February 2015, Volume 18, Issue 1, pp 63-70
[16] Hardyman W, B. A., Brown A, et al (2013). "*Mobile technology supporting trainee doctors' workplace learning and patient care: an evaluation.*" BMC Med Educ. 13(6).
[17] Mnkandla, E. (2009). "*About software engineering frameworks and methodologies*." AFRICON'09 IEEE.

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-368

A Method for Co-Designing Theory-Based Behaviour Change Systems for Health Promotion

Rebecka JANOLS[a,1] and Helena LINDGREN[b]
[a] *Department of Community Medicine and Rehabilitation, Umeå University, Sweden*
[b] *Department of Computing Science, Umeå University, Sweden*

Abstract. A methodology was defined and developed for designing theory-based behaviour change systems for health promotion that can be tailored to the individual. Theories from two research fields were combined with a participatory action research methodology. Two case studies applying the methodology were conducted. During and between group sessions the participants created material and designs following the behaviour change strategy themes, which were discussed, analysed and transformed into a design of a behaviour change system. Theories in behavioural change and persuasive technology guided the data collection, data analyses, and the design of a behaviour change system. The methodology has strong emphasis on the target group's participation in the design process. The different aspects brought forward related to behaviour change strategies defined in literature on persuasive technology, and the dynamics of these are associated to needs and motivation defined in literature on behaviour change. It was concluded that the methodology aids the integration of theories into a participatory action research design process, and aids the analyses and motivations of design choices.

Keywords: Health promotion, Participatory action research, behavioural change systems, persuasive design

1. Introduction

Behaviour change systems and persuasive technology aim at changing people's behaviour, typically towards a healthier behaviour [1]. A large number of behaviour change systems aim at increasing physical exercise or reducing behaviours such as smoking or alcohol intake. However, since behaviour change is very hard to accomplish, it is argued that theories on behaviour change are essential to apply in the process of designing and evaluating such systems [2, 3], which is rarely the case in existing examples. Some instruments and models incorporate some of the theories, such as the Behavioural Change Wheel [4], which is intended to function as instrument in the development process. However, another essential factor is how potential users can be engaged in the design process in order to increase ownership of the problem and instruments, and for capturing the diversity in attitudes, needs, preferences and motivation levels. This is essential in order to allow the system to be tailored to an

[1] Corresponding author, Department of Community Medicine and Rehabilitation, Umeå University, SE-90187 Umeå, Sweden; E-mail: rebecka.janols@umu.se.

individual's particular needs, preferences and situation. Such studies are rarely described in research literature, and participatory design methodologies do not incorporate knowledge about behaviour change.

Therefore, our aim is to define and develop a methodology for co-designing theory-based behaviour change systems for health promotion that can be tailored to the individual. The main research question is how theories on behaviour change and persuasive technology can be combined with participatory action design involving the targeted user group in order to i) capture participants' diverse attitudes, needs, preferences and motivation, and ii) design the adaptive behaviour necessary for supporting the change of behaviour in an individual. The main contribution of this research is a method that both allows users participate in the development, and that helps developers form designs of behaviour change systems that are based on theories on behaviour change and that adapts to the individual.

2. Methods

Research literature on behaviour change, behaviour change systems and persuasive technology was reviewed [3, 5-12]. Key concepts and models were identified and a selection was made of the concepts that were considered most important and relevant for the targeted domain. Aims and tasks to be conducted during the participatory action research design process were defined based on the theories, and formed an outline for the design process.

The outline was applied in two case studies, targeting respectively increase of desired behaviour and decrease of undesired behaviour. The first aimed at designing a behaviour change system for *increasing physical activity* among older adults. Since the role of the system was intended to be a kind of physical exercise advisor, the participants' relationship with health professionals and experiences relating to this was included in the process outline, and explored in the workshops. Workshops with nine older adults (divided into two groups), lead by two researchers, one designer and one physiotherapist, were conducted and audio recorded. Recordings, homework material and other material that was generated during workshops were analysed iteratively, and results were used for adapting the content of next workshops. The resulting design was analysed and evaluated based on theories on behavioural change and persuasive technology, and research on behavioural change systems. Based on the analyses the method was generalized and supplemented with additional themes. The generalized method was applied in a second case study. Two groups of designers and health professionals developed two initial designs of a *smoking cessation* intervention.

3. Results

The Participatory Action Research process based on theories on Behaviour Change and Persuasive technology is defined as a methodology (PAR-BCP) outlined Table 1. The outline functions as a checklist in the process of organizing tasks to be conducted during a design process. The methodology puts strong emphasis on the target group's participation in the design process. The following aspects are captured by applying the PAR-BCP Checklist, and explored together with the participants:

1. (diversity of) attitudes towards the targeted activity in focus and technology,

2. (diversity of) attitudes towards, and desires regarding the technology's potential pro-active behaviour such as encouraging and reminding messages,
3. (diversity of) attitudes, and desires regarding passive, summative communication of accomplishments,
4. (diversity of) attitudes, and desires regarding embedding social aspects and features, and
5. (diversity of) attitudes and desires regarding the system's potential agency and transparency.

These different aspects relate to the following behaviour change strategies defined in a framework of tailoring concepts [13]: *feedback, adaptation, inter-human interaction*, *user targeting*, *goal setting*, *context awareness* and *self-learning*. Different kinds of *feedback* are explicitly explored in the process, and the information about needs, motivation, diversity in attitudes, preferences etc., are elaborated for developing e.g., *goal setting* and *adaptation* strategies. The needs defined by Self-Determination Theory (SDT) are *competence*, *autonomy* and *relatedness* [7]. SDT distinguishes between *extrinsic* motivation, triggered by factors external to the individual, and *intrinsic* motivation, triggered by values internal to the individual. The system can be adapted based on an individual's level of motivation in relation to a particular activity.

The methodology incorporates the following seven design postulates for persuasive system design [3]: i) information technology is never neutral, ii) people like their views about the world to be organized and consistent, iii) direct and indirect routes are key persuasion strategies, iv) persuasion is often incremental, v) persuasion through persuasive systems should always be open/transparent. vi) persuasive systems should aim at unobtrusiveness and vii) persuasive systems should aim at being both useful and easy to use. These postulates are explored in the design process and embedded in the PAR-BCP Checklist (Table 1).

During and between group sessions the participants create material and designs following the behaviour change strategy themes, which are discussed, analysed and transformed into a design of a behaviour change system. The range and character of different attitudes and preferences illuminated in the design process can be captured and be taken into consideration when developing the adaptive, self-learning and context-awareness functionalities. These mechanisms are also important for supporting the incremental nature of persuasion.

3.1. Outcome of the Case Studies

The older adults in the first case study collaborated in creating new content and behaviour that was integrated in an existing mHealth application [14]. The second case study generated two different design proposals, where one provided solutions to obstacles found in the first study related to diversity in attitudes towards feedback.

In the first case study the initial version of the PAR-BCP Checklist was followed in detail with focus groups targeting each topic. In the second case study the PAR-BCP Checklist was used more as a checklist to assure that all aspects had been considered. The groups focused to large extent the nature of the smoking behaviour in order to identify what activates the behaviour in different individuals, and what are the perceived short term and long-term consequences. These were also the key topics in the meetings with the potential user group.

Both studies captured the diversity of attitudes towards the activity, the potential change of behaviour and towards using technology. The need for personalization was

illustrated by the very different activators that individuals mentioned in the smoking cessation case, and in the case with older adults by the opposite effects that the same message could have.

Table 1. PAR-BCP Outline: Generic outline of the PAR-process with exemplified activities.

Goal	Themes and Examples of Activities
Defining the behaviour / activity Understand and define the behaviour and activity to target.	**Define behaviour, and identify activators and consequences** Elaborate on what triggers the behaviour, and how consequences are perceived long term vs. short term.
Motivation level and attitudes Understand the target user group's/ participants' experiences and attitudes towards the targeted behaviour/activity and new technology.	**Motivation to conduct targeted behaviour / activity and use technology** Brainstorm about motivation and experiences of conducting the behaviour/activity. Participants express what they do/want to do with new technology in the activity context. Elaborate on *"Do I have to, do I want to, what's in it for me?"* long term vs. short term.
Unobtrusiveness, usefulness and ease of use Understand expectations and obstacles for use.	**Interaction design** Write a narrative about how to communicate through, and interact with (smart) technology when setting goals and when conducting the activities.
Pro-active support/feedback Build knowledge about what kind of feedback motivates participants to do the activity, and what feedback can be discouraging. Identify differences and similarities between individuals.	**Designing pro-active and passive feedback** Discuss different modalities for communication (images/text, sound, light, vibration etc). Write the messages they want to receive when they have or have not performed planned activity. Evaluate and rank feedback messages. Discussion about how they interpret and experience the content of the feedback messages.
Passive support/feedback Build knowledge about how to visualize activity progress and results / performance.	**Visualizing results and performance** Use cards/pictures for inspiration or sketch visualisations, discuss what the visualisations mean to each participant. Select and present favourites.
System's agency and transparency Explore what participants find important in the patient-health professional relationship, and in technology that builds trust.	**Agency and behaviour** Based on experiences with health professionals, describe desired behaviour and situations. Create profiles for agents/avatars with different personalities.
Inter-human interaction Explore how social interactions can support behaviour change.	**Designing for inter-human interaction** Describe desired social meeting points, and values to promote in these points that aid behaviour change.
Evaluating the design Evaluate design proposals.	**Evaluating the design** Evaluate design proposals in the form of mock-ups and prototypes, based on the above themes. Can be done during group sessions and/or in daily life.

The transparency of the system, i.e., knowing the purpose of its behaviour and having an image of who, or what, the sender of messages is, was shown to be very important for creating trust in the user. In the case of smoking cessation, the design choices converged towards making the system solely an instrument for the individual without interference from healthcare or health-based advice regarding smoking. Since the user was assumed to know well the health aspects of smoking that could form evidence-based rational arguments for cessation, and which have not so far made the persons quit smoking, such arguments were deliberately excluded. Instead, emotion-based arguments and goals associated to personal gains were included by the user. In the case with older adults they saw benefits of viewing the system as a coach with

physiotherapeutic expertise, but manifesting different attitudes to fit different persons' preferences.

4. Discussion and Conclusions

A method and an instrument are developed for the purpose of designing behaviour change systems that aim at having a sustainable effect on behaviour. The method fuses participatory action research methodology with theories and models of behaviour change and persuasive technology. The method and instrument were applied in two case studies with two different aims: reduction of unwanted behaviour, and increase of wanted behaviour. Comparison with other instruments, such as a behaviour model for persuasive design [12] and the behavioural change wheel [4] shows that the existing instruments do not guide how incorporating users in the design process.

The conclusion was made that the method aids the integration of theories into a participatory action research design process, and aids the analyses and motivations of design choices. As such the instrument can function as a checklist when designing behaviour change systems.

References

[1] Oinas-Kukkonen, H., A foundation for the study of behavior change support systems, *Personal Ubiquitous computing* 17 (2013), 1223–1235.
[2] Cowan, L.T., et al., Apps of steel: are exercise apps providing consumers with realistic expectations? A content analysis of exercise apps for presence of behavior change theory. *Health Education & Behavior* 40 (2012). 133-137.
[3] Oinas-Kukkonen, H. and M. Harjumaa, Persuasive Systems Design: Key Issues, Process Model, and System Features, *Communications of the Association for Information Systems* 28 (2009), 485-501.
[4] Michie, S., M.M. van Stralen, and R. West, The behaviour change wheel: a new method for characterising and designing behaviour change interventions, *Implementation Science* 6 (2011) 1-11.
[5] Matthews, J., K.T. Win, H. Oinas-Kukkonen, and M. Freeman, Persuasive Technology in Mobile Applications Promoting Physical Activity: a Systematic Review. *Journal of medical systems* 40 (2016) 1-13.
[6] Biddle, S.J.H., W. Brehm, M. Veheiden and M. Hopman-Rock, Population physical activity behaviour change: A review for the European College of Sport Science. *European Journal of Sport Science*, 12 (2012). 367-383.
[7] Ryan, R.M. and E.L. Deci, Self-Determination Theory and the Facilitation of Intrinsic Motivation, Social Development and Well-Being. *American Psychologist* 55 (2000) 68-78.
[8] Bandura A., Self-efficacy: toward a unifying theory of behavioral change, *Psychological review* 84 (1977) 191-215.
[9] Davis, R. et. al., Theories of behaviour and behaviour change across the social and behavioural sciences: a scoping review. *Health Psychology Review* 9 (2015) 323-344.
[10] Prochaksa, J.O. and C.C.D. Clemente, Transtheoretical therapy: Toward a more integrative model of change, *Psychotherapy: Theory, Research and Practice* 19 (1982) 276-288.
[11] Mohr, D.C., S.M. Schueller, E. Montague, M.N. Burns, and P. Rashidi, The behavioral intervention technology model: an integrated conceptual and technological framework for eHealth and mHealth interventions. *Journal of medical Internet research* 16 (2014) 146-152.
[12] Fogg, B.J. A behavior model for persuasive design, in Proceedings of the 4th international Conference on Persuasive Technology. 2009. ACM.
[13] Op den Akker, H., V.M. Jones, and H.J. Hermens, Tailoring real-time physical activity coaching systems: a literature survey and model, *User Model User-Adaption Interaction* 24 (2014), 351-392.
[14] Lindgren, H. L. Lundin-Olsson, P. Pohl, M., Sandlund. End Users Transforming Experiences into Formal Information and Process Models for Personalised Health Interventions, *Stud Health Technol Inform* 205 (2014), 378-382.

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-373

Persona Development and Educational Needs to Support Informal Caregivers

Zeina AL AWAR[a] and Craig KUZIEMSKY[a,1]
[a] *Telfer School of Management, University of Ottawa, Ottawa, ON, Canada*

Abstract. Informal caregivers are playing an increasing role in community based care delivery. Research is needed that looks at the educational needs of informal caregivers as a precursor to HIT design to support community care delivery. A challenge is informal caregivers have very diverse educational needs. Personas are an approach to describe user characteristics as part of systems design and this approach could be used to understand and categorize the various educational needs of informal caregivers. This paper addresses this research need and provides a method for persona development and the identification of educational needs for informal caregivers.

Keywords. Persona, informal caregiver, health education, tailoring

1. Introduction

As more care delivery is provided in community settings there is a need for the development of health information technology (HIT) to support care delivery outside traditional settings. The first step in designing HIT is identifying user requirements. The patient participatory medicine movement has led to an increase in applications designed to support patient led care delivery. Informal caregivers, such as family members or friends, are also playing an increasing role in community based care delivery. However, informal caregivers often struggle and feel insecure in their role [1]. Informal caregivers have substantial information needs as part of providing care but, to date, there is little research that has looked at the educational and information requirements of informal caregivers [11]. To overcome that shortcoming, we need interventions that support caregiver competence, preparedness, and confidence while providing care delivery [1]. These interventions must be based on evidence and designed to implement interventions for informal caregivers based on an assessment of their needs [2,3]. Informal caregivers have such vast needs that one size fits all solutions cannot be used but rather HIT solutions must be tailored to the specific needs of different types of informal caregivers. As a first step to HIT design, we need research that explores caregiver educational and care delivery needs.

While we cannot design HIT to support every context of a user group, we can design to support defined subsets of users. One such way of doing that is to develop user personas, described as "structured ways of typifying a group of users in text and pictorial formats."[4]. Personas describe user characteristics that go past mere demographics by capturing the mental processes users employ, including their

[1] Corresponding author, E-mail: kuziemsky@telfer.uottawa.ca

expectations, prior experiences, and projected behaviour [4]. A persona is typically based on a narrative that specifies the goals that a fictional but representative user has for using an artifact, while also describing their environment, qualification, attitudes, and the tasks they need to accomplish.

While personas are a valuable way to define and shape requirements for end users, there is a need for papers that present approaches for developing personas [5]. A particular need is for persona developers to understand the various educational needs of informal caregivers [5]. This paper addresses this research need and provides a method for identifying education needs and developing personas to support the delivery of educational needs to informal caregivers.

2. Materials and Methods

2.1. Data Sources

Two data sources informed the study. First, we conducted 8 semi-structured interviews with eight informal caregivers. Three of the interviews were in person and five were conducted via telephone. Five of the participants were female and three were male; their ages ranged from 38 to 70 years. All were either related or married to their patients. The questions were designed to capture the experiences of the participants and then use them to derive their unexpressed needs.

Second, a nursing focus group was also conducted with four nurses with experience caring for palliative cancer patients in a home setting. The focus group protocol was also semi-structured and open-ended, and the participants discussed their experiences caring for palliative patients and dealing with their informal caregivers.

2.2. Data Analysis

Qualitative content analysis focused on coding and extracting themes from the data collected in the in-depth interviews. Our analysis had two goals. First was to develop user personas. Participant characteristics were laid out in table format and used to derive user personas using Miles, Huberman, and Saldaña's recommendations for making inferences from data matrices [6]. Characteristics extracted from the interviews to determine the individual background and circumstances of each participant were mapped out on a spreadsheet. The spreadsheet was then analyzed for patterns that identify the differences in their caregiving needs following a set of recommendations for making inferences from matrices, including observing patterns and themes, making comparisons and contrasts, clustering, and making and verifying explicit conclusions [6].

Our second goal was to understand the educational needs of informal caregivers to identify how they differed according to the personas. Educational needs were identified using the Needs Assessment design [7], which articulates user experiences as a prerequisite to identifying their needs. We studied caregiver needs according to three categories of needs from the Needs Assessment design [7]: Take-It-for-Granted Needs, which are the minimum expectations of a service, Typical Wants, which are attributes that end-users watch for and assess, and Attractive Needs, which are those that the end-user is usually completely unaware of having.

3. Results

3.1. Caregiver Personas

Persona characteristics extracted from the interviews were mapped out on a spreadsheet to determine the individual background and circumstances of each participant. Upon examining the caregiving habits and the expressed needs of the interview participants, no difference was noted among different ages, genders, geographical settings, or educational levels. Furthermore, the level of comfort with computer technology did not seem to affect their information-seeking habits, as every one of the participants used technology to search for information related to their caregiving role at one point during their care delivery. From our data analysis we identified the level of illness of the patient being cared for and the level of caregiving intensity as two key dimensions to structure personas.

The patient's illness level directly influences their functional status and their care needs, often dictating the information needs of the caregiver [8]. The second version of the Palliative Performance Scale (PPS) [9] was used to estimate the functional and performance level of patients cared for by the interview participants. PPS assigns a score in percent increments of 10 with 100% being full ambulation and self-care, 60% having reduced ambulation and some assistance need, and 20% being bed bound with total caregiving needed. Transcripts were re-examined and a new column was added to the spreadsheet to assign a PPS score to each patient. We grouped patients into 3 PPS categories – 50-60%, 30-40% and 10-20%, as each category represents distinct caregiving needs. People with a PPS score of 70% and above are fully functional and do not require a caregiver.

Caregiver intensity was defined using the values set by Jacobs, Laporte, Van Houtven, & Coyte 2014 [10]. A new column was added to the spreadsheet to assign a category to each participant. Participants were considered less-intense caregivers if they provided less than 5 hours of care per week, mid-intensity if they provided 5-15 hours of care per week, and high intensity if they provided 15 or more hours of care each week [10].

Table 1. Five persona categories according to caregiving intensity (vertical axis) and Palliative Performance Scale Version 2 (PPSv2) (horizontal axis)

PPS / Caregiving intensity	10-20%	30-40%	50-60%
<5 hours/week			LILN
5-15 hours/week	MIHN	MIIN	MILN
>15 hours/week	HIAN		

Table 1 shows the five categories of personas we developed: Less-intense low-need (LILN) caregiver, Medium-intensity low-need (MILN) caregiver, Medium-intensity intermediate-need (MIIN) caregiver, Medium-intensity high-need (MIHN) caregiver, and High-intensity All needs (HIAN) caregiver. Those with a score of 40% and below all needed a caregiver with a level of involvement greater than 5 hours per week. Only one high intensity category (HIAN) was created as high-intensity

caregivers were defined by one role where the caregiver would take care of all patient care needs. This is especially prevalent among caregivers of unstable patients, who could be high need one day and intermediate or low need the next day.

3.2. Informal Caregiver Educational Needs

In the second part of our analysis we identified informal caregiver educational needs according to four categories: Presentation, implementation, practical caregiving and information. The categories started through deductive codes [e.g. 12-13] that were then extended through inductive codes from our data analysis. Presentation refers to how information is presented to the end-users. Implementation referred to the way information and recommendations can be incorporated into the daily life of the end-users and the policies of the health care system under which they live. Practical caregiving is the caregiving components that deal with patient care issues including emergency measures, equipment, medication management, mobility, nutrition, pain, and physical symptoms. Information deals with the clinical information that informal caregivers need to understand including different aspects of the patient's illness, treatment options, available resources, and complications to be expected such as functional decline. Information on functional decline was one of the key attractive (i.e. unexpressed) needs from participants.

3.3. Differences in Caregiving Needs by Personas

After determination of the informal caregiver needs as discussed in the previous section, the next task was to discover how the five personas differed from each other with respect to caregiver needs. While the categories of caregiver needs were consistent across the personas, the operationalization of needs differed. The biggest difference among the different personas was in the practical caregiving category. When patients transition from low into medium and high intensity categories, their medication regimen increased significantly, and thus medication management (e.g. medication logs and information on drug-drug interactions) became a priority need. Further, as caregiving transitions from low into medium intensity, functional decline of the patient becomes more significant and information (e.g. symptom onset and management) and tools (e.g. equipment) to understand and support a patient in decline becomes a priority need. The different caregiving needs and how they must be presented to support different personas become the basis for HIT requirements. For example, as patients become less mobile in the medium and high intensity personas, videos and other instructional tools for supporting mobility and transfers would be an HIT requirement.

4. Discussion

This study presented an approach for developing caregiving personas as a means of identifying and structuring educational needs as a precursor for HIT design. Our approach extends existing work on personas in two ways. First, while existing research on personas has been based on one dimension (i.e. geographic) our personas are based on two dimensions (caregiving intensity and the functional and performance level of the patient). Second, our personas emphasize different caregiving experiences and the ways that caregiver needs changed over time. One novel feature of our persona

development approach is that it allows for the determination of caregiver needs within a persona at a specific period of a patient's illness, but it also supports the descending transition into subsequent personas with additional and/or more sophisticated needs than the previous persona.

Our next step is to use the personas and information needs to derive requirements for the design and evaluation of a prototype HIT to support informal caregivers. The educational needs we identified (e.g. information, practical caregiving) will be converted to system design requirements. A limitation of this study is that we have only developed personas for one very specific group of caregivers. Testing our personas and caregiver needs in other settings and caregiver groups is also future work to be done.

Acknowledgements

We acknowledge funding support from a research grant from the Telfer School of Management and a Discovery grant from the Natural Sciences and Engineering Research Council of Canada.

References

[1] E. Harrop, A. Byrne, A. Nelson, "It's alright to ask for help": findings from a qualitative study exploring the information and support needs of family carers at the end of life, *BMC Palliative Care,* **13(22)** (2014)

[2] L.L. Northouse, M.C. Katapodi, A.M. Schafenacker, D. Weiss, D, The impact of caregiving on the psychological well-being of family caregivers and cancer patients, *Seminars in Oncology Nursing,* **28(4)** (2012), 236-245.

[3] J. McDonald, E. McKinlay, S. Keeling, W. Levack, Becoming an expert carer: the process of family carers learning to manage technical health procedures at home, *Journal of Advanced Nursing*, **72 (9)** (2016), 2173-2184.

[4] C. LeRouge, J. Ma, S. Sneha, K. Tolle, K, User profiles and personas in the design and development of consumer health technologies. *International Journal of Medical Informatics,* **82(11)** (2013), e251-e268.

[5] S. Vosbergen, J.M.R. Mulder-Wiggers, J.P. Lacroix, H.M.C. Kemps, R.A. Kraaijenhagen, M.W.M. Jaspers, N. Peek, Using personas to tailor educational messages to the preferences of coronary heart disease patients, *Journal of Biomedical Informatics*, **53** (2015), 100-112.

[6] M.B. Miles, A.M. Huberman, J. Saldaña, *Qualitative Data Analysis: A Methods Sourcebook* (3 ed.). SAGE Publications, Inc, 2014

[7] Y. Akao., Quality Function Deployment: Integrating Customer Requirements into Product Design, Productivity Press, Cambridge, Massachusetts., 1990

[8] S.B. Wackerbarth, M.M. Johnson, Essential information and support needs of family caregivers. *Patient Education & Counseling, **47*(2)** (2002), 95-100.

[9] Palliative Performance Scale (PPSv2) version 2. In G. M. Downing, & W. Wainwright (Eds.), *Medical Care of the Dying* (4th ed., p. 121). Victoria, BC: Victoria Hospice Society., 2006

[10] J. C. Jacobs, A. Laporte, C. H.Van Houtven, P.C. Coyte, Caregiving intensity and retirement status in Canada, *Social Science & Medicine,* **102** (2014), 74-82.

[11] C. Levine, D. Halper, A. Peist, D. A. Gould, Bridging troubled waters: Family caregivers, transitions, and long-term care, *Health Affairs,* **29(1)** (2010), 116-124.

[12] N. Nijland, J. van Gemert-Pijnen H. Boer, M. Steehouder, E. Seydel E, Evaluation of Internet-Based Technology for Supporting Self-Care: Problems Encountered by Patients and Caregivers When Using Self-Care Applications, *J Med Internet Res,* **10(2)** (2008), e13.

[13] K.T. Washington, S.E. Meadows, S.G. Elliott, R.J. Koopman, Information needs of informal caregivers of older adults with chronic health conditions. *Patient Education and Counseling,* **83** (2001), 37-44.

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-378

An Approach for Enhancing Adoption, Use and Utility of Shared Digital Health Records in Rural Australian Communities

Helen ALMOND[a,b1] , Elizabeth CUMMINGS[b] , Paul TURNER[a]
[a]*School of Engineering & ICT, University of Tasmania*
[b]*School of Health Sciences, University of Tasmania*

Abstract. Internationally, shared digital health records are considered an important addition to improving modern health care provision. Australia launched its version, My Health Record (MyHR), in 2012 but has experienced low adoption and challenges in practical implementation and evaluation. Individuals living with complex and chronic conditions in rural and remote communities often experience challenges in obtaining equitable access to health care provision. They are also supposed to face additional barriers to adopting and using eHealth services. This paper reports on research investigating adoption, use and utility of MyHR, in rural remote Australian community settings. Based on this research an approach for improving national roll out of MyHR is presented. The approach highlights a means to understand and engage communities with complex care needs, to support their adoption and use of digital tools. It also draws attention to holistic methods for evaluating and assessing impact at individual, community and health care provision levels.

Keywords. My Health Record, rural communities, digital complex care pathways.

1. Introduction

Internationally, shared digital health records are being acknowledged as fundamental in modern health care provision. Governments and policy makers are responding to the challenges; recognising adoption of individual shared digital health records will improve integration thereby delivering improved quality and efficiencies. The support required by individuals to access their health data electronically is an approach, which appears to be being underestimated [1-5]. The use of an individual shared digital health record can offer a continuum of equitable health care provision in disease prevention, management, treatment and reduction in disparities in care [3, 6].

Those individuals who could benefit most from shared digital health records are those who create the largest burden on health care delivery; people with complex chronic conditions, living in rural remote communities, disadvantaged because of their limited access to efficient, quality health care provision [1, 3].

Australia launched its shared digital health record, My Health Record (MyHR), in 2012; it remains without significant adoption or any evident holistic implementation

[1] Corresponding author: Helen Almond, School of School of Engineering & ICT, University of Tasmania. E-mail Helen.Almond@utas.edu.au

and evaluation framework. This paper reports on research investigating the adoption, use and utility of MyHR, in rural remote Australian community settings. Based on this research, an approach for improving national roll out of MyHR is presented. This approach highlights a means to understand and engage individuals and communities with complex care needs, to support their adoption and use of digital tools and draws attention to holistic methods for evaluating and assessing effects at individual, community and health care provision levels. The contribution being to support future successful implementation and uptake of similar projects.

2. Method

Individuals involved in a rural remote lifestyle modification program expressed an interest in recording and reflecting their lifestyle goals and modification progress, initiated the eHealth research project. The tool for documenting the intervention was the Personal Health Notes, a designated section within MyHR.

Based on a participatory philosophy, a qualitative community based participatory eHealth research project was developed and conducted in southern Tasmanian rural remote community health care settings [7]. Once ethical approval was granted (HREC-H0013781), 21 research partners, 15 females and six males, 40-89 years, with two or more complex chronic conditions, were recruited from three rural remote settings.

Data were collected in three phases: pre-adoption, adoption and post-adoption of MyHR. Data collection tools were: groups convened in the three rural remote settings, held in each phase, the partners adopting and initiating use of MyHR during phase two, and partner semi-structured interviews held during phase one and three. All encounters were audio recorded and transcribed verbatim. Group memos, MyHR personal health notes, a reflective diary and project documentation provided additional evidence.

Data analysis was structured using a systematic, flexible thematic framework approach. This was achieved in three stages: data management, data description and data explanation [8]. Data transcription ran concurrently with data collection, which assisted confirmation of transcripts, by individual partners and the groups and ensured a comprehensive, consistent in-depth systematic data analysis.

3. Results

Three key findings emerged, which support a dynamic conceptualisation of MyHR.

3.1. Individual, Community and Societal Understanding is Inherent

For an individual with complex chronic conditions equitable, person-centred, integrated health care provision is fundamental and can be enabled by and embedded in MyHR, irrespective of physical or cognitive ability. The research partners agreed there is a need for a flexible individualised digital health care record. They want a role, ownership of and responsibility for their health and health information. Those who need MyHR most, need the most support; MyHR should not exclude anyone. There is an inherent desire to have MyHR where they can engage and play a meaningful role in

defining and communicating their: identity, complex chronic conditions, and main problem, as part of their digital complex care pathways.

> two individuals may have the same diagnosis, but how it affects them and where they go from there can be totally different…what I try to get through to a lot of people … is I know my body and how they affect me. If [MyHR] allows me to communicate that … It doesn't need to be just doctors. It's our bodies and our lives that this is all about and the two should be working together… but that tends to get lost (CH23).

3.2. Individual, Community Skills are Recognised and Valued as Shared Experiences.

Using MyHR lead to increased knowledge, competence and confidence and a changed perception of computer use. As the research partners became experienced they evolved, finding new ways of using MyHR. When a diverse group of individuals was offered the opportunity to engage in a collaborative, coordinated, community project, they became familiar with the use of MyHR, issues of design and utility emerged. When rural communities are involved and supported through adoption and initial use of MyHR, they become engaged, innovative and knowledgeable providing an irreplaceable source of user information and experience.

> this research is great, everybody likes to be involved and feel they have an opportunity to give … We're starting learning together … learning computer together you share so much instead of individuals wondering how do I get back to it…working with buddies… you always learn by how somebody uses something. That's been really good and the research has tapped into the local [groups] and all these initiatives around. Yes if you want to know how anything is going to work you give it to the users (CH13).

3.3. Rural Communities Need to be Included in all Digital Health Care Interactions

The research partners viewed and valued MyHR, as a digital health care tool that can provide multiple opportunities for the delivery of their health care. The research partners learned they were not alone; they had MyHR and complex chronic condition experiences acknowledged and confirmed by others. Once the partners were empowered to explore MyHR they learned and shared the value of it as potential for engagement and an access point to wider health care provision.

> things are getting done, people can find out about more of their diseases … it's got to be an improvement to find out these things. It's an opportunity especially a rural community like this. It helps, it gets people together, they discuss things and it makes it a lot better (CH6).

In 2012 MyHR espoused an opportunity, for those who wished, to 'opt-in', the ability to add and receive agreed health information. However, pre-adoption of MyHR, those research partners who had heard of MyHR were unsure of its purpose and none had registered. When the community identified an opportunity to experience MyHR, every research partner engaged, immediately increasing their knowledge and opportunity to communicate and access their health information. Without reservation, all research partners, irrespective of their ability, described the benefits MyHR would have on their future engagement in their digital complex care pathways. Some partners found having access to MyHR useful and rewarding, regardless of the level of provider involvement. Others communicated frustration because of the lack of engagement by their health care providers [9].

Although access to health care provision may have improved, through the sharing of resource experiences and provision of a community lifestyle modification program, shared access to information prepared by health care providers, via MyHR, was not

possible. The research did not find the community, disinterested or disengaged because of limited access to efficient, quality health care provision. However, their health care providers were reported being dismissive.

> I tried to get my Drs involved, … my specialist …I gave them access when I went for my appointment and I talked to [them] about it. [They were] saying [they] didn't like the fact that other doctors could see basically, or other people could see. If my other Dr had what was it, access to it then he could see my [disease] stuff and stuff like that. It wasn't a problem to me but it's a problem to [them] (O3).

4. Discussion

Everyday living and working with complex chronic conditions requires understanding and acknowledgment of their multifocal complex non-linear care requirements [10]. In context, these individuals and their carers, living in a rural community, a complex adaptive model of care can assist by conceptualising MyHR, as one component embedded within a digital complex care pathway, which equitably engages in an integrated, person-centred approach to health care provision [11]. Using the characteristics of complex adaptive system theory: *agents who learn*, individuals can and will learn and react to changes in information; *interconnection*, changes in patterns of interactions, among agents and introducing; *self-organisation*, order can be created without explicit hierarchical direction; and *coevolution*, the system and the environment influence each others development, to the research findings provides a conceptual framework, which can support the meaningful, adoption use and utility of MyHR within digital complex care environments [12].

Embedding MyHR within all digital complex care pathways may assist in; the adoption use and utility of MyHR, for individuals with complex chronic conditions, and all health care providers in their transition from linear, episodic, medico-centric models of health care provision, toward achieving equitable, person-centred, integrated health care provision.

The function of any digital health care system is to deliver improvements in health care provision and experience; new eHealth technology should be evaluated to ensure accountability and continued improvement [13]. This requires looking at how MyHR can help and spending time and effort to review process to achieve the ultimate goals: efficiency and quality. When implementing the relatively new digital health record solution, MyHR, it is important to separate health care provision, which needs to be maintained and enhanced, from process that needs to be revised to include questions and explanations as to why we do the things we do and how could we do things better? [1-4]. Including complex adaptive system theory, within an adoption, use and utility of MyHR framework, for all stakeholders involved in the delivery of digital complex care pathways could be instrumental in answering these questions.

5. Conclusion

The adoption of MyHR will take place nationally but for use and utility to be realised regionally and locally, communities require engagement, information and support, or critical mass will not be achieved [14, 15]. The research demonstrates the importance of listening to all stakeholder experiences to help shape future strategies, which are

responsive to the diversity of illness, individual and community experience. Valuing the expertise of people with complex chronic conditions empowers them to adopt interactive and cooperative relationships with their health care providers. Understanding the individual's experiences is key in the delivery of digital, equitable person-centred care, integrated care.

During the research, the community learned to overcome issues of technophobia and identified a need for infrastructure, which supports the adoption and use of MyHR, highlighting the necessity to identify and educate all health care providers regarding; their continued resistance to shared digital health information, and the professional use of information and computing technologies. There is a requirement of communities and individuals to be engaged and valued during their transition through their complex chronic conditions supplemented by MyHR.

This research provided the impetus for a community to commit to and engage with MyHR. In Australia and internationally a digital complex care framework should; engage individuals and communities with complex care needs, and be incorporated to provide a holistic approach toward future roll out, evaluation and assessment of impact at individual, community and health care provision levels.

References

[1] S. Whetton, *Health informatics: a socio-technical perspective*. Oxford University Press, 2005.

[2] Project Integrate, in *Benchmarking Integrated Care for better Management of Chronic and Age-related Conditions in Europe.* (2016), vol.2016.

[3] eHealth. Initiative, A study and report on the use of ehealth tools for chronic disease care among socially disadvantaged populations, Final Report. California, 2012.

[4] T. Greenhalgh, H. Potts, G. Wong, P. Bark, D. Swinglehurst, Tensions and Paradoxes in Electronic Patient Record Research: A Systematic Literature Review Using the Meta-narrative Method, *Milbank Quarterly* **87** (2009), 729-788.

[5] C. Nohr, M. C. Wong, P. Turner, H. Almond, L. Parv, H. Gilstad, S. Koch, G. A. Harethardottir, H. Hypponen, R. Marcilly, A. Sheik, K. Day, A. Kushniruk, Citizens' Access to Their Digital Health Data in Eleven Countries - A Comparative Study. *Stud Health Technol Inform* **228** (2016), 685-689.

[6] E. Murray, J. Burns, T. See, R. Lai, I. Nazareth, Interactive Health Communication Applications for people with chronic disease. *The Cochrane database of systematic reviews*, CD004274, 2004.

[7] B. Israel, A. Schulz, E. Parker, A. Becker, Review of community-based research: assessing partnership approaches to improve public health. *Annual review of public health* **19** (1998), 173-202.

[8] J. Ritchie, L. Spencer, in *Analyzing Qualitative Data,* A. Bryman, R. Burgess, Eds. Routledge, London, 1994, 173-194.

[9] H. Almond, E. Cummings, P. Turner, Avoiding Failure for Australia's Digital Health Record: The Findings from a Rural eHealth Participatory Research Project. *Studies in health technology and informatics* **227** (2016), 8-13.

[10] K. Lorig, H. Holman, Self-management education: history, definition, outcomes, and mechanisms. *Annals of behavioral medicine: a publication of the Society of Behavioral Medicine* **26** (2003),1-7

[11] WHO, "Framework on integrated, people-centred health services," *Sixty-Ninth World Health Assembly Provisional agenda item 16.1,* 2016.

[12] L. Leykum, J. Pugh, V. Lawrence, M. Parchman, P. Noel, J. Cornell, R. McDaniel, Jr., Organizational interventions employing principles of complexity science have improved outcomes for patients with Type II diabetes. *Implement Sci* **2** (2007), 28.

[13] C. Martin, J. Sturmberg, Complex adaptive chronic care. *J Eval Clin Pract* **15** (2009), 571-577.

[14] T. Greenhalgh, K. Stramer, T. Bratan, E. Byrne, J. Russell, S. Hinder, H. Potts, The Devil's in the Detail: Final report of the independent evaluation of the Summary Care Record and Health Space programmes, University College London, 2010.

[15] R. Jolly, The E Health Revolution: Easier Said Than Done. Canberra: Parliamentary Library, 2011.

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-383

Exploring Innovation Capabilities of Hospital CIOs: An Empirical Assessment

Moritz ESDAR[a,1], Jan-David LIEBE[a], Jan-Patrick WEIß[a], Ursula HÜBNER[a]
[a]*Health Informatics Research Group, Osnabrück University AS, Germany*

Abstract. Hospital CIOs play a central role in the adoption of innovative health IT. Until now, it remained unclear which particular conditions constitute their capability to innovate in terms of intrapersonal as well as organisational factors. An inventory of 20 items was developed to capture these conditions and examined by analysing data obtained from 164 German hospital CIOs. Principal component analysis resulted in three internally consistent components that constitute large portions of the CIOs innovation capability: organisational innovation culture, entrepreneurship personality and openness towards users. Results were used to build composite indicators that allow further evaluations.

Keywords. Innovation capability, Innovation management, composite indicator, hospital CIOs

1. Introduction

A rich body of studies agrees that Chief Information Officers (CIOs) occupy a central position in visioning, guiding and implementing IT based innovations [1,2]. These innovations can generally be defined as changes of products and processes that result from the adoption of IT and are new to the given organisation [3]. In the hospital context, IT innovations mostly fall under the category of process innovations (e.g. the widespread implementation of a new clinical decision support system or telemedicine solutions) that lead to significant changes of the related workflows or process outcomes [4].

Even though empirical investigations could substantiate the critical role of CIOs to foster IT innovations in the industrial sector [e.g. 5], there is no scientific evidence about the innovation capability of CIOs in healthcare, particularly in hospitals. In fact, there are reasons to assume, that hospital CIOs innovation attempts might be challenged by specific social and organisational circumstances [6]. Although medical decision-making processes cannot be entirely automated, as they require complex medical knowledge as well as the clinician's individual experience [7], the respective workflows can still be significantly improved by providing accurate data and information. The goal hereby is to seamlessly integrate the information flow into the clinician's work practice and particularly support advanced clinical processes. This phenomenon is described by the information logistics construct [8] which matches one of the criteria for innovation proposed by Hübner [4]. At this point, the innovational capability of the CIO often makes the difference between IT success and failure as they not only have to be very considerate

[1] Corresponding author, Moritz Esdar, Osnabrück University of AS, Health Informatics Research Group, PO Box 1940, 49009 Osnabrück, Germany; E-mail: m.esdar@hs-osnabrueck.de.

with the clinician's expectations, autonomy and the peculiarities of the medical workflows, but also act in an environment that is characterised by financial restrictions [2]. Health information technology (HIT) is known to be frequently perceived as a mere cost factor by the executive board and therefore often lacks adequate support [9]. Specific innovation capabilities of CIOs may therefore be constituted by their ability to mediate between highly skilled professions and to act as an enabler within a potentially restrictive organisational environment. This is also referred to as *intrapreneurship* [10].

Up to date, empirical studies about hospital CIOs mainly focus on questions related to their structural power (position, reporting level etc.) [11] and on how these factors correlate with given CIO roles or decision types [9]. Whereas these approaches are meaningful in themselves, they often neglect the underlying personality (e.g. the CIOs views and attitudes) and environmental patterns (e.g. the executive board's attitude towards IT). Our goal, therefore, was to 1) shed light in what constitutes innovation capabilities of hospital CIOs both in terms of intrapersonal as well as organisational factors and 2) determine how the innovation capability construct can be operationalised.

2. Methods

Original scales were developed based on Patterson and colleagues' [3] framework of people relevant resources for innovation in organisations that distinguishes environmental factors tied to the workplace (external dimension) and intrapersonal factors (internal dimension). We initially operationalised each domain by 40 items on different types of scales. Pre-testing the inventory (undertaken by 6 hospital CIOs and 8 health IT researchers) resulted in a final inventory of 20 Items, 10 for each domain measured by Likert scales. Data were collected between February and April 2016 via an online survey. We obtained 164 valid responses from a total of 1284 contacted German CIOs (response rate 12.77%).

In order to 1) explore underlying patterns of our data, 2) reduce the inventory to a set of variables that describe innovation capability, 3) test the discriminant and convergent validity and reliability (using Cronbach's alpha) as well as to 4) develop an empirically founded composite indicator, we performed principal component analysis (PCA) [12]. Following strong recommendations of the methodological literature [13], we applied the underlying variable (UV) approach using polychoric correlation coefficients since all included variables were measured on ordinal scales. Applicability of the correlation matrix was evaluated based on the Kaiser-Meyer-Olkin (KMO) criterion and Bartlett's test of sphericity. Components were extracted if their eigenvalue exceeded 1, if all components explained at least 50% of the total variance and based on consulting the scree plot. We allowed the extracted components to correlate by using oblique rotation since we did not assume them to be entirely distinct from each other. To obtain a set of meaningful and discriminant items, we gradually removed items that could not be fitted in the component structure (i.e. showed heavy cross loadings or component loadings $< .5$ across different model solutions). The final solution was tested for reliability and then interpreted in a group discussion of eight experts (comprising health IT scientists, statisticians, management researchers and a psychologist).

Component loadings and eigenvalues were used to deploy a weighting scheme adapted from the Organisation for Economic Co-operation and Development (OECD) [14] in order to build a composite indicator for each component and for the full inventory that accentuates the components and corrects for statistically overlapping information.

3. Results

According to a KMO measure of .73 and a significant result of Bartlett's test of sphericity our data proved to be suited for PCA. Moreover, the sample to variable ratio was 13:1 and therefore was above recommended minimum ratios which typically range between 5:1 to 10:1 [15]. In the course of reducing the inventory, we attained a final set of 13 items that were ideally reflected in a solution comprising 3 components (Table 1) explaining 51% of the total variance. Interrelations between the components remained low with correlation coefficients less than .15.

Table 1. Component loading matrix. Loading below .3 are left blank

Item	Component		
	1	2	3
"Our executive board actively promotes innovative IT solutions."	.82		
"Our hospital has a well-defined future vision that is also being pursued by the IT department."	.74		
"Our hospital shows great flexibility when it comes to employing innovative IT."	.74		
"Our hospital is way too rigid on all levels of hierarchy to employ IT in a strategically meaningful fashion." (reverse coded)	.70		
"IT is perceived as a mere expense factor by our executive board way too often" (reverse coded)	.68		
"Our IT department is only able to provide highly valuable services if every employee consistently covers an unchanged range of tasks" (reverse coded)		.68	
"My work mainly consists of realising the wishes and ideas of other people." (reverse coded)		.66	
"As the person in charge of IT, I first of all rely on well-established IT solutions." (reverse coded)		.57	
"My work motivation would be significantly higher if I was paid adequately to my knowledge and skills." (reverse coded)		.52	
"A CIO has to first of all take care of technical and not people issues." (reverse coded)			.76
"It is very important to me to have great knowledge of the clinical processes in our hospital."			.63
"Listening and giving advice are the core competencies in my role as a CIO."			.62
"It is very important to us to incorporate the different clinical end users in our IT projects."			.56

The full scale showed acceptable reliability in terms of internal consistency with $\alpha = .71$. Similarly, component 1 showed good internal consistency ($\alpha = .78$) whereas components 2 ($\alpha = .64$) and 3 ($\alpha = .52$) showed lower but acceptable reliability values given the relatively low number of associated items. The components were interpreted as "organisational innovation culture" (component 1), "entrepreneurship personality" (component 2) and "openness towards users" (component 3).

Table 2. Descriptive statistics of the developed composite indicators (n = 164)

Composite Indicator	Mean	SD	Range	Skewness	Kurtosis
Full inventory	55.86	12.29	59.67	.15	-.34
Component 1	53.33	20.54	100	-.18	-.04
Component 2	42.25	15.23	86.66	.17	.41
Component 3	74.98	14.27	67.06	-.35	.06

Table 2 displays the distributional properties of the calculated composite indicators that were built using the data driven weighting scheme referred to above. Each indicator was scaled to range between 0 (complete disagreement with all related statements) and 100 (complete agreement with all related statements)

4. Discussion

The importance of the CIOs' innovation capability increases with the growing potentials and diffusion of HIT. Hitherto it remained unclear which particular conditions constitute these capabilities (research question 1) and how these conditions can be operationalised (research question 2).

Results of the PCA and subsequent score development indicated two essential findings with regard to question 1. At first, it confirmed a clear empirical distinction of the external dimension opposed to internal (intrapersonal) aspects, as all items of component 1 were originally intended to measure the environmental dimension. In contrast to interpreting this component as the general organisational environment it can be specified as organisational innovation culture and support from the executive board. This aligns well with existing theoretical knowledge pointing out the importance of top management support [16] that gives HIT based innovations the required flexibility [17], active financial promotion, and guiding principles and vision [2] for innovative HIT to prosper. All these aspects seem to be indicative of a coherent dimension describing a fundamentally positive attitude towards innovative IT within the organisation. The second finding reveals that the previously assumed "internal dimension" has to be broken down into two separate dimensions, i.e. into "entrepreneurship personality" and "openness towards users". "Entrepreneurship personality" is a composition of traits that embraces intrinsic motivation and self-determination, a mindset of internal freedom to deviate from established paths and to take risks. This is a clear contrast to tayloristic attitudes. "Openness towards users" is a trait that is closely related with "involvement of users" and "participation" of users, which is a well-known success factor in systems engineering [5] and in innovation alike [8]. Our initial thoughts on CIOs' specific requirement of closely incorporating the clinician's interests when striving for HIT innovations now show an empirical manifestation in this component.

With regard to question 2, the analysis led to a full set of 13 items measuring three different dimensions of innovation capability. Whereas internal consistency measures where satisfying for component 1, reliability measures for component 2 and 3 were marginally acceptable. Greater precision and redundancy in these domains are desirable in further investigations. However, the full set of items showed an acceptable internal consistency with $\alpha = .71$. It was reduced on the grounds of the PCA results. Although this is a common methodical approach [12], it potentially threatens the construct's integrity since a few aspects were removed which might have been retained if they were captured with greater redundancy (i.e. more questions). It therefore is reasonable to assume that there might be more to innovation capability beyond our model's dimensions. Another limitation arises from the modest response rate of 12.77% that might have caused a non-response bias in our sample. The results can therefore only be generalised with caution and require further validation in different samples.

The resulting composite indicator is normally distributed around a mean of 56 points (out of 100). Thus, innovation capability seems to be moderately advanced in German hospitals with clear potential for development. It is most notably that component 3 "openness towards users" showed significantly higher values with $\bar{x} = 75$ whereas component 2 "entrepreneurship personality" only showed an average score of 42. Many hospital CIOs apparently understand the importance of participation and user focus but are still surprisingly prone to a work approach that does not create much space for self-determination and deviation from established paths. The actual impact of the composite indicator and its subscales still needs to be tested against innovation performance

measures to further assess their validity and to determine which particular aspects most strongly drive HIT innovations. This study provides a fundamental toolset to do so.

5. Conclusion

This study gives insight into the constituents of the construct innovation capability of CIOs and defines a set of items to operationalise this construct. In contrast to previous findings, we not only distinguish between internal and environmental factors, but clearly denote them specifying the dimensions unique to hospital CIOs. We hereby lay the foundation of a psychometric inventory to measure innovation capability.

References

[1] M. Broadbent & E.S. Kitzis, *The New CIO Leader, Harv. Bus. School Press*, 2005.
[2] D.E. Leidner, Preston, D., & Chen, D. An examination of the antecedents and consequences of organizational IT innovation in hospitals, *J. Strateg. Inf. Syst*. **19** (2010), 154–170.
[3] F. Patterson, M.Kerrin & G.R. Geraldine, Characteristics and behaviours of innovative people in organisations, *Literature Review prepared for the NESTA Policy & Research Unit* (2009), 1–63.
[4] U. Hübner, What Are Complex eHealth Innovations and How Do You Measure Them?, *Methods Inf. Med.* **54** (2015), 319–327.
[5] S. Watts & J.C. Henderson, Innovative IT climates: CIO perspectives, *J. Strateg. Inf. Syst*. **15** (2006), 125–151.
[6] K. Cresswell & A. Sheikh, Organizational issues in the implementation and adoption of health information technology innovations: an interpretative review, *Int. J. Med. Inform.* **82** (2013), e73-86.
[7] R. Lenz & M. Reichert, IT support for healthcare processes – premises, challenges, perspectives, *Data. Knowl. Eng.* **61** (2007), 39–58.
[8] M. Esdar, U. Hübner, J.D. Liebe, J. Hüsers & J. Thye, Understanding latent structures of clinical information logistics: A bottom-up approach for model building and validating the workflow composite score, *Int. J. Med. Inform.* **97** (2017), 210–220
[9] F. Köbler, J. Fähling, H. Krcmar & J.M. Leimeister, IT governance and types of IT decision makers in German hospitals, *Bus. Inf. Syst. Eng.* **2** (2010), 359–370.
[10] K.L. Heinze & K. Weber, Toward Organizational Pluralism. Institutional Intrapreneurship in Integrative Medicine, *Organ. Sci.* **27** (2015), 157–172.
[11] D. Burke, N. Menachemi & R. Brooks, Health care CIOs: assessing their fit in the organizational hierarchy and their influence on information technology capability, *Health Care Manag.* **25** (2006), 167–172.
[12] G.O. Otieno, T. Hinako, A. Motohiro, K. Daisuke & N. Keiko, Measuring effectiveness of electronic medical records systems: towards building a composite index for benchmarking hospitals, *Int. J. Med. Inform.* **77** (2008), 657–669.
[13] T.A. Brown, *Confirmatory Factor Analysis for Applied Research*, The Guilford Press, New York, 2015.
[14] M. Nardo, M. Saisana, A. Saltelli, S. Tarantola, A. Hoffmann & E. Giovannini, *Handbook on constructing composite indicators: Methodology and user guide*, OECD Paris, 2008.
[15] R.C. MacCallum, K.F. Widaman, S. Zhang & S. Hong, *Sample size in factor analysis*, Psychol. Methods **4** (1999), 84–99.
[16] D.H. Smaltz, V. Sambamurthy & R. Agarwal, The antecedents of CIO role effectiveness in Organizations. An empirical study in the healthcare sector, *IEEE Trans. Eng. Manage.* **53** (2006), 207–222.
[17] R.V. Bradley, T.A. Byrd, J.L. Pridmore, R. Thrasher, R.M.E. Pratt & V.W.A. Mbarika, An empirical examination of antecedents and consequences of IT governance in US hospitals, *J. Inf. Technol.* **27** (2012), 156–177.

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-388

The Invisibility of Disadvantage: Why Do We Not Notice?

Chris SHOWELL[a,1], Elizabeth CUMMINGS[b] and Paul TURNER[a]
[a] *eHealth Services Research Group, University of Tasmania, Australia*
[b] *Discipline of Nursing and Midwifery, University of Tasmania, Australia*

Abstract. Personal health records (PHRs) offer tantalising benefits for patients and healthcare providers, including improvements in patient-provider communication, patient empowerment, and access to data and information. A suspicion that disadvantaged patients are less likely to use or benefit from PHRs stimulated a research agenda that included: (a) a literature review; and (b) empirical analysis of eight years' hospital admission and discharge data linked to measures of patient social disadvantage. The results demonstrated an association between disadvantage, increased use of public hospital services and barriers to PHR use. These findings may appear self-evident, but dramatically highlight how disadvantaged patients continue to be overlooked in many e-health design processes, and are rarely a focus of user centred design. The paper concludes by briefly considering the implications of this invisibility.

Keywords. Personal health records, Barriers, Disadvantage.

1. Introduction

Current trends in patient centred health informatics suggest that there are significant benefits for patients when they make use of personal health record systems (PHRs), which have been defined as "...a private, secure application through which an individual may access, manage, and share his or her health information." [1, p. 244] The potential benefits include better patient-provider communication, patient empowerment, access to health self-management, and improved access to data and information [2], [3]. However there is little evidence that the incorporation of PHRs into the everyday provision of health care will be an appropriate solution for all patients. As Rigby and Ammenwerth have recently noted, the development and use of informatics in health care has been marked by a "…lack of clear strategic investment decisions, and lack of evidence, [which] interlink." [4, p. 4]

This paper summarises and contextualises research completed as part of a doctoral thesis, based on empirical data analysis of eight years' hospital admission and discharge data linked to measures of patient social disadvantage [5]. The research originated in a lingering concern that patients who are at a socioeconomic disadvantage are likely to face a 'triple threat' because of the difficulties they face with low income, increased health needs, and challenges with textual, technical, and health literacy.

[1] Chris Showell, eHealth Services Research Group, School of Engineering & ICT, Faculty of Science Engineering and Technology, PB 65 University of Tasmania, Sandy Bay 7001 TAS Australia. Email: chris.showell@utas.edu.au

These individuals are often 'disempowered, disengaged and disconnected', and have been largely invisible in the process of health records development. [6], [7]

The research was conducted in two phases. The first phase involved a critical appraisal of the research literature to characterise the relationship between socioeconomic disadvantage and personal electronic health records. In this phase refereed publications were used to delineate, test and validate ideas and concepts. These publications considered: the way in which ordinary literate citizens might be marginalised by the use within PHRs of specialised medical language and terminology, SNOMED CT in particular [8]; the omission of these same citizens from discussions about the direction of Australia's ehealth policy [9]; and a widespread tendency for PHRs to be designed by and for an insider elite of 'People Like Us', with the approaches taken in the design, implementation and evaluation of PHRs being likely to ignore the preferences, needs and capabilities of disadvantaged users [6], [7].

Publications from the first phase of the research identified aspects of personal electronic health record systems that could limit their usefulness for disadvantaged patients, and prompted the subsequent research described below. This was conducted during 2013 and 2014, using a two-part approach. First, a literature review explored current evidence about barriers to the uptake and continued use of personal electronic health records. Second, empirical data analysis was used in an attempt to delineate, in an Australian setting, a group of disadvantaged healthcare users likely to face higher barriers to the adoption and use of PHRs.

2. Methods

The literature review targeted publications providing evidence about barriers which might prevent the adoption of a PHR, or interfere with its continued use. The review covered refereed items published in English after 2003. Publications which focused on barriers for providers or provider organisations were excluded. Thematic analysis was used to identify core themes within publications

The investigation of disadvantage, health service use and PHR barriers involved an empirical analysis of data covering the 96 geographic areas in Tasmania. This data included: three of the Australian Bureau of Statistics' Socio Economic Indicators for Areas (as measures of disadvantage); deidentified patient records for 800,000 public hospital admitted episodes and 1.3 million emergency department attendances (around eight years' data); and area measures of education, internet access and qualifications (as proxy measures for text- , technical- and health literacy). Cluster analysis of measures for disadvantage, healthcare use and identified PHR barriers was used to identify groups within the data, with choropleth maps used to visualise significant trends.

3. Results

The literature review [10] found evidence of a range of barriers that may interfere with the adoption and continued use of PHRs. Across the 40 included publications there were 21 individual barriers identified. These included: age; race or ethnicity; income and socioeconomic status; education; text, technical and health literacy; internet and computer access; and disability. Twelve of the 21 barriers identified had an association

with socioeconomic disadvantage. Barriers were found in all phases of PHR adoption, and in all types of investigation. As a secondary outcome, the review also identified a number of PHR evaluations that may have introduced a selection bias by actively excluding low capability participants. Since it was not possible to deduce the relative importance of particular PHR barriers from the frequency of their appearance in the research literature, the review did not attempt to rank the significance or prevalence of the barriers identified.

The analysis of hospital service use [11] showed that patients from areas with a low socioeconomic status used public hospital services at a higher rate, and had longer inpatient stays; these areas also had a higher incidence of factors associated with barriers to PHR use. Cluster analysis identified two distinct subgroups of areas with disadvantaged, low capability users receiving more public hospital care, and proxy measures suggesting barriers to PHR adoption and use, in contrast to more privileged capable subgroups using much less care. Figure 1 presents choropleth maps which highlight the dichotomy between these clusters.

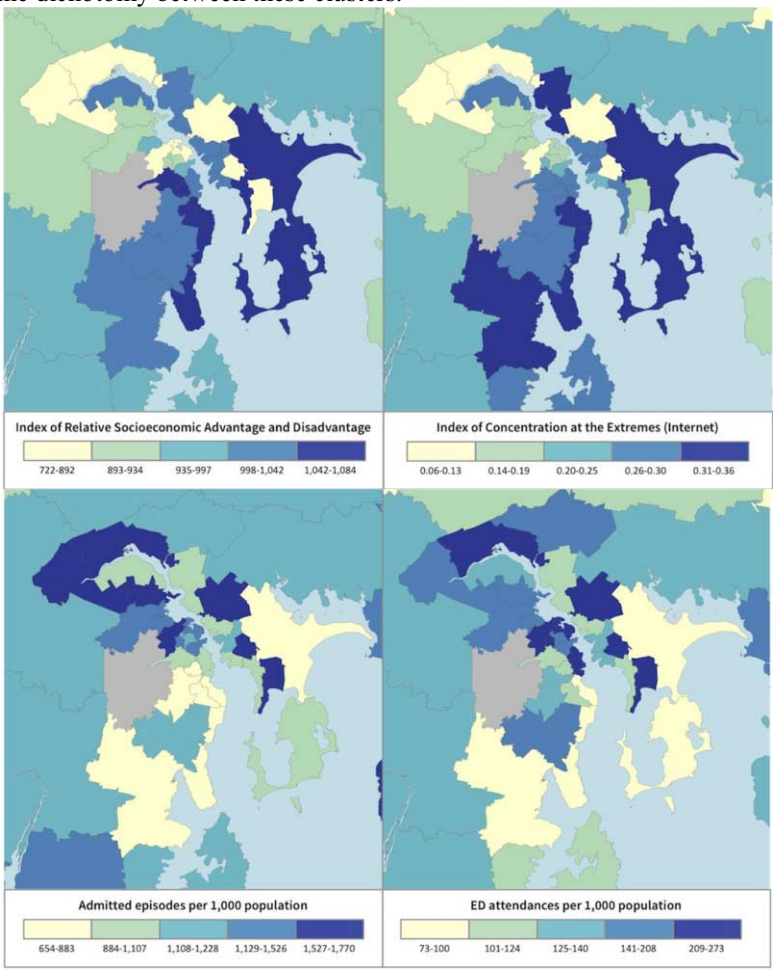


Figure 1. Maps of increasing financial resources and internet use (top) and hospital and ED use.

The maps focus on the most populous area of Tasmania, with Hobart, the state capital, in the middle of each map, and the Derwent estuary to the bottom right. The upper two maps show measures of overall financial wellbeing (L) and internet access (R). Those in the second row show admitted hospital episodes (L) and emergency department visits (R), both measured per 1,000 population. Colours change from lighter to darker as values increase, with significant overlap between the palest neighbourhoods in the first row (more disadvantage and worse internet access) and the darkest neighbourhoods in the second (more health service use and PHR barrier factors).

4. Discussion

This paper has briefly summarised detailed evidence identifying those areas in Tasmania whose populations have higher levels of socioeconomic disadvantage, higher use of public hospital services, and proxy measures suggestive of lower literacy. These areas also have a higher incidence of chronic disease and of capability barriers which are likely to limit any potential benefit from PHR use. PHRs as they are currently implemented are unlikely to provide a universal solution for problems with healthcare delivery or communication. These findings highlight a need for more attention to be paid to the implications of disadvantage during PHR design, implementation, and evaluation. A careful assessment is required of the relevance of each potential barrier within the population being considered as end users. From a health informatics perspective, this paper argues that those involved in the design, implementation and evaluation of personal electronic health record systems appear to be overlooking a crucial requirement for such PHR systems – that they be "fit for purpose" in the context of their intended use and intended end-users.

None of this is new, but this research dramatically highlights how disadvantaged patients continue to be marginalised in many e-health design processes, and are rarely a focus of user centred design. More specifically this research highlights that:

- Current approaches to research on personal electronic health records mean that the socially disadvantaged are invisible, often being discounted as 'non-adopters' or 'not qualifying' for participation in relevant PHR studies. Without special attention, PHRs will continue to ignore the 'disempowered, disengaged and disconnected';
- Patients from disadvantaged neighbourhoods in Tasmania use public hospital services to a greater extent than those from privileged neighbourhoods, and display characteristics which are indicative of barriers to PHR use;
- There is a risk that disadvantaged patients will receive worse healthcare as a result of a focus on personal electronic health records as currently realised. Diversion of health resources to the implementation of PHRs may result in an increased inequity in healthcare outcomes and contribute to a growing e-health divide; and
- User centred design for PHRs may help to address this issue, but only with the participation of a truly representative group of potential users, including the disadvantaged.

5. Conclusion

The findings of this research confirm what we already know – poor people have poor health, use more health services, and are less likely to benefit from PHRs. What is remarkable is how the disadvantaged have remained invisible in the development of e-health systems. When these results were shared with healthcare professionals who provide direct care to patients, a common response was "so what?" They saw the findings as so obvious that they did not warrant further commentary or action. This paper argues that this is simply not a tenable response for e-health professionals; we must not ignore these barriers in the design and configuration of health information systems intended for direct use by patients, or for healthcare services more generally.

Given that individuals experiencing some form of socioeconomic disadvantage represent 10% to 15% of the overall population in most Western societies, and a larger proportion of the demand on healthcare services, it is worrying that there seem to be so few recorded attempts to holistically tailor the design of PHRs to be suitable for patients who cannot read, have difficulty using technology, and struggle with the interpretation of health concepts and terminology. Just as perplexing is the observation that few patient healthcare services are specifically tailored for these same individuals.

Policymakers, informaticians, health service managers and healthcare providers need to look carefully behind the cloak of invisibility that marginalises any consideration of the implications of disadvantage in our technology initiatives. This is a crucial way of ensuring that any benefits generated by ehealth innovation are being shared equitably. If we do not, then those whose healthcare is most in need of improvement will be the least likely to benefit from having (at least theoretical) access to a PHRs, and our e-health initiatives may end up contributing to, rather than ameliorating the social divides that already pervade our societies.

References

[1] D. A. Jones, J. P. Shipman, D. A. Plaut, and C. R. Selden, "Characteristics of personal health records: findings of the Medical Library Association/National Library of Medicine Joint Electronic Personal Health Record Task Force," *J. Med. Libr. Assoc. JMLA*, vol. 98, no. 3, pp. 243–249, Jul. 2010.

[2] P. C. Tang and D. Lansky, "The missing link: bridging the patient-provider health information gap," *Health Aff. Proj. Hope*, vol. 24, no. 5, pp. 1290–1295, Oct. 2005.

[3] C. Pagliari, D. Detmer, and P. Singleton, "Potential of electronic personal health records," *BMJ*, vol. 335, no. 7615, pp. 330–333, Aug. 2007.

[4] M. Rigby and E. Ammenwerth, "The Need for Evidence in Health Informatics," *Stud. Health Technol. Inform.*, vol. 222, pp. 3–13, 2016.

[5] C. Showell, "Hidden in plain sight: personal health records and the invisibility cloak of disadvantage," University of Tasmania, 2014.

[6] C. Showell and P. Turner, "The PLU problem: are we designing personal ehealth for people like us?," *Stud. Health Technol. Inform.*, vol. 183, pp. 276–280, Feb. 2013.

[7] C. Showell and P. Turner, "Personal health records are designed for people like us," *Stud. Health Technol. Inform.*, vol. 192, p. 1037, Nov. 2013.

[8] C. Showell, E. Cummings, and P. Turner, "Language games and patient-centred eHealth," *Stud. Health Technol. Inform.*, vol. 155, pp. 55–61, 2010.

[9] C. Showell, "Citizens, patients and policy: a challenge for Australia's electronic health record," *Health Inf. Manag. J.*, vol. 40, no. 2, pp. 39–43, Jun. 2011.

[10] C. Showell, "Barriers to the use of personal health records by patients: A structured review," (Under review), 2016.

[11] C. Showell, "Disadvantage and public hospital use in Tasmania," presented at the Sustainable Healthcare Conference, Hobart, Tasmania, 2015.

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-393

Using Healthcare Work Process Modelling in Hospitals to Increase the Fit Between the Healthcare Workflow and the Electronic Medical Record

David MORQUIN[a,b], Roxana OLOGEANU-TADDEI[a,1], Ludivine WATBLED [c]
[a] *Montpellier Research in Management, University of Montpellier, France*
[b] *University Hospital of Montpellier, France*
[c] *University Hospital of Lille, INSERM CIC-IT 1403, France*

Abstract. As with other organizations, hospitals tend to promote unrealistic expectations related to software implementaton. Quite often the real issue is a misfit between the software and organizational factors. Our paper shows how work process modelling within the hospital can reduce this misfit according to the vision developped by the ergonomics and the management of information systems. This idea is supported by two cases in two different University Hospitals in France, in which using work process modelling lead to identification of problems and their causes, and solutions. Modelling requires time, which may be considered costly by senior hospital managers, but also should be considered as an investment in order to achieve expected goals.

Keywords. Modelization, Business process reengineering, Information System management, Humans Factors

1. Introduction

Technological innovations are often shrouded in myths surrounding their ability to solve social and economic problems [1] or they are the cause of "magical thinking" that buying or providing a tool leads automatically to its adoption and use in line with the expected benefits [2]. Change management, adaptation of technology and generally the underlying investment in human and financial resources are often under looked [3]. As with other organizations, hospitals promote unrealistic expectations without taking into account human and organizational factors. Thus, Electronic Medical Records (EMR) are supposed to have various benefits for the medical practices such as providing easy access to patient record documentation and increasing the accuracy of these records [4], reducing potential medical errors and improving the quality of patient care [5]. In addition, authors have shown that uses are not always optimal as designed by managers or editors and may lead to workarounds [6], increasing data recording time and increasing risk for errors. One of the main explanations of this fact consists in there being a misfit between the healthcare workflow process and the workflow process imposed by

[1] Corresponding author, Montpellier Research Management, EA 4557, University of Montpellier, F-34090 Montpellier, France; E-mail: roxana.ologeanu-taddei@umontpellier.fr

the technology [7,8]. This misfit is related to the multiplicity and the diversity of clinical paths and of clinician practices and needs between the medical departments [9], and the lack of coordination [10] among sequential and interdependent actors [11].
Consequently, we will show in this paper that it is possible to increase the fit between the healthcare workflow and the EMR as well as with related tools, by identifying of the causes of the problems related to these misfits and, consequently, of different solutions, with the shareholders involved. We will show also that the workflow process modelling method helps to make this diagnosis and to select solutions based on a shared representation of the work process.

2. Background

Workflow process modelling is usually used in three different frameworks. The first framework is that of IT, including medical informatics. Modelling is understood as developing a model that could be used for IT development. Healthcare information models are generally formal specifications for representing the healthcare content structure and semantics within electronic health record systems in order to implement and evaluate these models [12,13]. The second framework is the Business Process Rengineering (BPR) using information technology as a support to enable a complete redesign of work processes [14] in order to achive business ojectives as reducing the cost or the delay or increasing quality. Thus, the modelling process is the support of their analysis; it aims to identify, capture and characterize the components of the process via various tools to obtain a representation or mapping [15]. The third framework is that of ergonomics and analysis of the work, which focuses on the operator, the task, work, and the work context into which the operators, their work and their tasks are inserted [16,17]. The model is a representation of the worflow or work process defined as embedded process logic that is resourced by an appropriate set of human and technical resources [18], shared by the different operators concerned. In this paper, we consider work process modelling within the hospital according to the second and the third frameworks. Tasks and work are considered synonimous, related to the the work done, without a prescriptive nor normative representation of this work.

3. Methodology

Our method, which we call Healthcare Work Processes Modeling in Hospitals (HWPMH) is characterized by the following steps: 1) Modeling is done only for each specific care process or department, and not as a “one-size-fits-all” method for all the departments or hospital; 2) Modelling is focused on the work including informational workflow (kind of information, support); 3) Modelling is done at the request of an operator or of a researcher, after identifying a problem related to the misfit beetween the organization and the technology; 4) Modelling, as description of the work proccess, leads to a shared representation of the work with the operators, which is mapped by a schematic, using BPMN softwares (Business process modelling notation) BonitaSoft© and Bizagy© ; 5) This representation is presented during a meeting with the actors concerned (operators and other stakeholders as Human Ressources managers,

Information Technology specialists, healthcare professionals), which have to accept or modify the proposed solutions.

This method was implemented simultaneously and independently in two different French University Hospitals (Montpellier and Lille), by an employee of each hospital – an ergonomist psychologist from the ISD in Lille and a physician of the Delegation to the Hospital Information in Montpellier. Both individuals are also researchers. In the two cases, the modelling method was the same. The main difference consisted in the method used to collect data in order to model the work process (which consisted in a group interview with the operator for the Montpellier Hospital and in observation for the Lille Hospital). During the meetings with operators or other actors concerned verbatims were transcribed.

The specific processes modeled are presented in the figures 1 and 2.

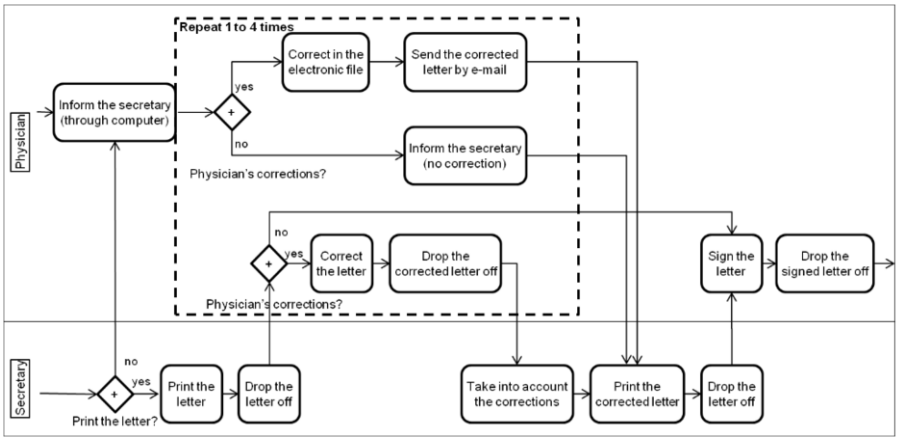


Figure 1. Detail of the final process of transmission of the discharge letter in Lille Hospital

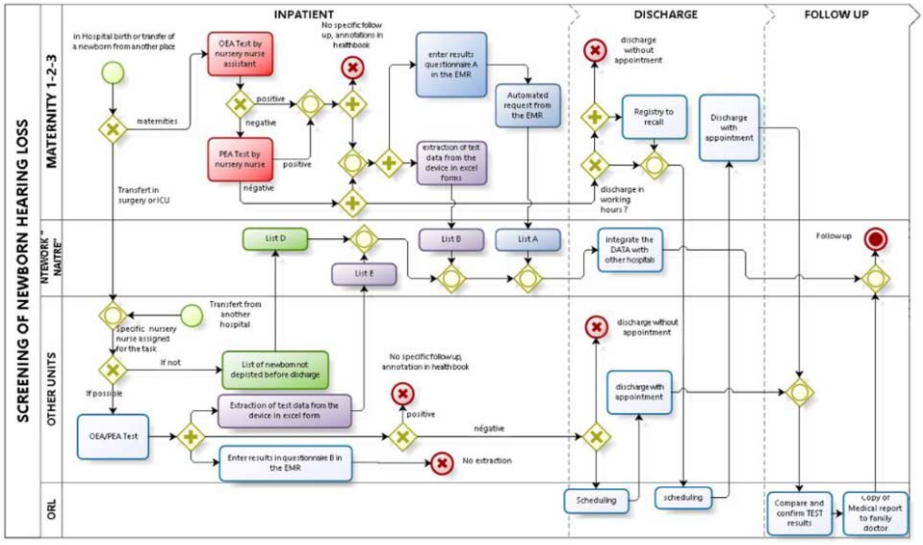


Figure 2. Process of information exchange for screening of newborn hearing in Montpellier Hospital

4. Results

Our results are the outputs of the HWPMH method. Several meetings with operators and other stakeholders triggered a final model as share representation as the work process related to the EMR use. The table below synthesises the inputs or problems of the modelling method, their causes (highlighted by HWPMH) and their outputs or solutions.

Table 1. HWPMH Inputs and Outputs (Problems, Causes and Solutions)

	Lille	**Montpellier**
Input (problem)	Decreasing time in discharge letter transmission related to implementation of digital dictation is not effectively as expected	Lack of completeness of the hearing screening for newborn in the hospital (<80%) : missing data for the Regional Monitoring Network
Inputs (Causes)	Digital dictation is just a part of the process of producing and sending the discharge letter. Before sending, the validation step depends on two intermediate steps : the number of reviewers and of the recording of the letter as a pdf file, which is related to the secretary's work tasks and workload.	Different recording sources relates to different care sectors for newborns Lack of accuracy of the data from the screening of newborn hearing into the EMR. Different pathways to provide data (differents forms in the EMR while data are extracted from the measuring devices also) Different way of scheduling appointments with: otorhinolaryngologists according to the release time (no possibility for the discharge during the night or weekend)
Outputs (Solutions proposed)	Decrease the number of reviewers (no more than 2) Send the discharge letter on the day of discharge or in advance.	Accurate definition of the different terms and data for newborn hearing screening Design and implement an unique form, integrated into the EMR, including automated data exportation and alert system when the fields remains empty after 3 days or at discharge. Optimize the scheduling system to be used by nursery nurses and nursery assistants to take specific appointment with otorhinolaryngologists.

5. Discussion

While the methods used in the the two hospitals to gather information for workflow process modelling were different, schematics resulting from the modelling were designed in a similar way. In both cases, modelling as description of the work led to identification of the causes of the problems and, consequently, of different solutions. The representation of the work using schematics was shared by all the operators and actors concerned. This is a condition of the success of our method because coordination among interdependent actors [11] is generally missing. Both cases overall showed that beyond technology organizational changes are requested in order to reduce misfit and to increase benefits expected from the implementation of the software. Modelling requires time, which may be considered costly by hospital top managers while it should be considered as un investment in order to achieve expected goals. Furthemore, workflow process modelling provides better understanding of the fit between organization and technology, through information as mediating concept: information as designed and performed by work on the one side and designed and provided by technology on the other side. Information is not a specific work or task related to a specific occupation but is the core

of healthcare work and that of healthcare software. The accuracy of data and information recorded has to match information requested for the daily healthcare work in order to perform tasks into an interdependent, sequentiel and transverse processes.

6. Conclusion

Evidence from two French hospitals shows how work process modelling based on consensus allows identification of problems, of their causes and solutions. Thus, we suggest that hospitals have to choose between the IT "magic bullet" as a charming way to represent quick benefits from techonology implementation without organizational efforts and the workflow process modelling as a pragmatic way to increase fit between the organization and the technology requesting time and patience.

References

[1] V. Scardigli, Le sens de la technique, Collection Sociologie d'aujourd'hui., PUF, 1992.
[2] M.L. Markus, and R.I. Benjamin, The Magic Bullet Theory in IT-Enabled Transformation, *Sloan Manage. Rev.* **38** (1997) 55–68.
[3] G. Pare, Implementing clinical information systems: a multiple-case study within a US hospital, *Heal. Serv. Manag. Res.* **15** (2002) 71–92.
[4] J.S. Ash, D.F. Sittig, E.G. Poon, K. Guappone, E. Campbell, and R.H. Dykstra, The Extent and Importance of Unintended Consequences Related to Computerized Provider Order Entry, *J. Am. Med. Informatics Assoc.* **14** (2007) 415–423.
[5] D.W. Bates, and A.A. Gawande, Improving safety with information technology, *N Engl J Med.* **348** (2003) 2526–2534.
[6] S.Y. Park, S.Y. Lee, and Y. Chen, The effects of EMR deployment on doctors' work practices: A qualitative study in the emergency department of a teaching hospital, *Int. J. Med. Inform.* **81** (2012) 204–217.
[7] K. Cresswell, and A. Sheikh, Organizational issues in the implementation and adoption of health information technology innovations: An interpretative review, *Int. J. Med. Inform.* **82** (2013) e73–e86.
[8] L. Lamothe, La reconfiguration des hôpitaux un défi d'ordre professionnel, *Ruptures, Rev. Transdiscipl. En Santé*. **6** (1999) 132–148.
[9] C. Pascal, La gestion par processus à l'hôpital entre procédure et création de valeur, *Rev. Française Gest.* **146** (2003) 191–204.
[10] S. Glouberman, and H. Mintzberg, Managing the care of health and the cure of disease: arguments for the importance of integration., *Health Care Manage. Rev.* **26** (2001) 85-87-90.
[11] J.D. Thompson, Organizations in action: Social science bases of administrative theory, Transaction publishers, 2014.
[12] S. Medlock, J.C. Wyatt, V.L. Patel, E.H. Shortliffe, and A. Abu-Hanna, Modeling information flows in clinical decision support: key insights for enhancing system effectiveness, *J Am Med Inf. Assoc*. **23** (2016) 1001–1006.
[13] A. Moreno-Conde, and T. Austin, Evaluation of clinical information modeling tools, *J.* (2016)
[14] T.H. Davenport, and J.E. Sbort, The New Industrial Engineering: Information Technology and Business Process Redesign, *MIT Sloan Manag. Rev.* **31** (1990) 11–27.
[15] R.S. Aguilar-Savén, Business process modelling: Review and framework, *Int. J. Prod. Econ.* **90** (2004) 129–149.
[16] J. Leplat, L'analyse du travail en psychologie ergonomique, Octarès, 1992.
[17] H. Chaudet, F. Anceaux, M.C. Beuscart, S. Pelayo, and L. Pellegrin, Facteurs humains et ergonomie en informatique médicale, *Inform. Médicale, E-Santé—fondements Appl. New York Springer*. (2013) 495–520.
[18] K. A. Chatha, J.O. Ajaefobi, and R.H. Weston, Enriched multi-process modelling in support of the life cycle engineering of Business Processes, *Int. J. Prod. Res.* **45** (2007) 103–141.

This page intentionally left blank

4. Knowledge Management

This page intentionally left blank

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-401

Reference Architecture Model Enabling Standards Interoperability

Bernd BLOBEL[a,b,1]
[a] *Medical Faculty, University of Regensburg, Germany*
[b] *eHealth Competence Center Bavaria, Deggendorf Institute of Technology, Germany*

Abstract. Advanced health and social services paradigms are supported by a comprehensive set of domains managed by different scientific disciplines. Interoperability has to evolve beyond information and communication technology (ICT) concerns, including the real world business domains and their processes, but also the individual context of all actors involved. So, the system must properly reflect the environment in front and around the computer as essential and even defining part of the health system. This paper introduces an ICT-independent system-theoretical, ontology-driven reference architecture model allowing the representation and harmonization of all domains involved including the transformation into an appropriate ICT design and implementation. The entire process is completely formalized and can therefore be fully automated.

Keywords. Interoperability; system; ontology harmonization; interoperability reference architecture; framework

1. Introduction

1.1 Interoperability Issues

Interoperability as defined by IEEE as "ability of two or more systems or components to exchange information and to use the information that has been exchanged" must be able to interconnect those in the business case involved systems technically by guaranteeing signal and protocol compatibility (technical interop.). With growing knowledge about the business case shared among the involved components in advance or at runtime, information systems interoperability evolves in following steps:

1. Sharing data about the business case at different level via simple electronic data interchange (EDI) (syntactic interop.) or structured messaging like HL7v2 (syntactic interop.);
2. Information sharing enabling the understanding of underlying concepts of the business case represented in openEHR Archetypes, EN/ISO 13940 ContSys concepts, HL7v3 artifacts, or HL7 FHIR® resources (semantic interop.);
3. Taking the information and communication technology (ICT) part related actions for realizing the business objectives (service interop.).

All the aforementioned interoperability levels happen between ICT components supporting the business case, while the business case is usually a non-ICT but real

[1] Corresponding Author. Bernd Blobel, PhD, FACMI, FACHI, FHL7, FEFMI, Professor; Medical Faculty, University of Regensburg, Germany; Email: bernd.blobel@klinik.uni-regensburg.de

world scenario. According to the Good Modeling Practice, the relevant stakeholders define the provided view of the model as well as the way of structuring and naming the concepts of the problem space. First capturing key concepts and key relations at a high level of abstraction, different abstraction levels should be used iteratively, where the first iteration is performed in a top-down manner to guarantee the conceptual integrity of the model. This requires meeting design principles such as orthogonality, generality, parsimony, and propriety [1]. Therefore, the IEEE interoperability definition has to be extended to include the domains involved in the business case resulting in cross-domain cooperation, which requires sharing and harmonizing those domains' knowledge represented by related ontologies and policies. As the domain experts do not share equal levels of experiences, training, knowledge, etc., shortly summarized in skills, the final interoperability level is skills-based interoperability resulting in moderated individualized end-user collaboration based on individual ontologies. In other words, interoperability turns from technical aspects of ICT and data sharing to domain-specific aspects of policy, knowledge and skills harmonization. In the lack of administrative power as well as a priori shared and enforced knowledge and policies, missing interoperability pre-requisites must be distributed and harmonized at runtime by advanced interoperability solutions.

The advancement of medicine from an empirical, phenomenological approach towards systems medicine enabling personalized, preventive, predictive, participative precision medicine for individually tailored care requires the cross-disciplinary understanding of the status of the individual and its context as well as its correct, consistent and formal representation for integration in the ICT system as part of the business system. Harmonizing and sharing all instances of individual cases and their informational representations in advance is impossible, as a) nobody can pre-define what will happen, and b) sharing all thinkable instances will lead to a representation system's complexity which is un-determined and not manageable.

The non-ICT interoperability is not just defining the ICT interoperability solution, but is the real challenge in the game. In most interop-standards specifications, both facts are ignored.

1.2 Standards and Specification Issues

Meanwhile, many standards and specifications have been developed by different Standards Development Organizations (SDOs) to enable cooperation between actors in health business cases. Thereby, specifications usually address different aspects of the business from a specific domain's perspective, considered by domain experts using their specific methodologies as well as their specific terminologies, at best based on related domain ontologies. In real world business systems, usually different domains are involved, so requiring an integration of those different specifications, thereby growing regarding the complexity and multi-disciplinary characteristics when moving to more advanced healthcare paradigms.

For harmonizing domain or use case specific specifications when adding a new specification or changing/extending the business case, especially when including another domain, currently a mutual adaptation and harmonization is performed, resulting in a revision process of the impacted standards and specifications. With increasing complexity and variability of the system and the diversity of its subsystems and components, the lifetime of domain specific specifications goes down.

The alternative way of a priori harmonizing the aforementioned highly complex, highly dynamic, multi-disciplinary/multi-domain advanced healthcare system by representing it by one domain's terminology/ontology or - even worse - by using ICT ontologies fails. The same holds when using one domain's representational style and models or standards as reference or master that all other domains and their experts must adhere to, e.g., by enforcing biologists, physicians, philosophers and artists to think and represent in UML and the 78 concepts of the ICT base ontology [2].

Therefore, an adaptive approach is required to sharing and harmonizing ICT, domain, and personal ontologies and conditions at runtime.

2. Methods

For meeting the non-ICT interoperability challenges, an abstract domain-independent representation of systems is deployed, based on a system-theoretical, architecture-centric, ontology-driven approach [3, 4]. The mathematical concept representation using the universal type theory in combination with systems engineering methodologies allows representing any system architecturally (i.e. the system's components, their functions and internal as well as external relations) by generically describing its composition/decomposition as well as the aspects (domains) of the system relevant in a specific context (e.g. business case), instantiated using those domains' ontologies. The reference architecture model - here focus on the business domain - can be used recursively, so representing, e.g., the real-world systems' continuum from elementary particles to the universe (Figure 1).

Additionally to agreeing on one or more, and at best standards-based, ICT ontologies, the agreed domains' knowledge, but also individual (language, education, skills, experiences, social and psychological aspects, etc.) and environmental context must be represented, harmonized and communicated by instantiating the system's architectural components and behavior through the domain-specific ontologies and policies.

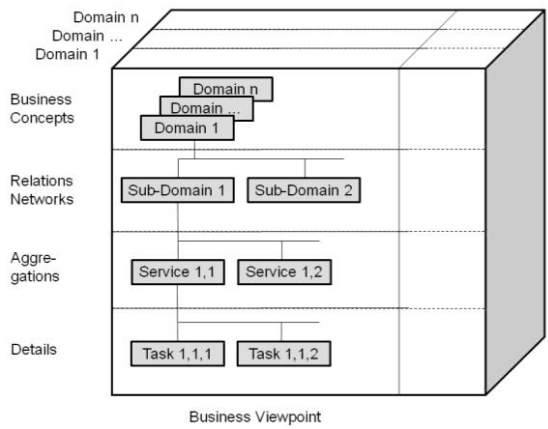


Figure 1. Granularity Levels of the Interoperability Reference Architecture Model

By combining that model with ISO/IEC 10746 RM-ODP, the Interoperability Reference Architecture Model (introduced in the nineties as Generic Component Model

- GCM) as well as the applicable rules - the Interoperability Reference Architecture Model Framework - (also known as GCM Framework) is completed [4, 5].

3. Results

This reference architecture model allows consistently describing any complex real world system's structure and behavior by representing concepts and relationships of the domain-specific sub-system at the real world system component's level of granularity using the specific domain ontologies. In other words, the domain-specific subsystem (e.g. a domain-specific standard or specification) is not harmonized any more by peer-to-peer and case-by-case revisions (Figure 2a), but is re-engineered using the Interoperability Reference Architecture Model (Figure 2b). Bound to the GCM Framework, inter-domain relationships must happen at the same level of granularity [5]. To get there, intra-domain specializations/generalizations have to be performed. In summary, the Interoperability Reference Architecture Model supports ontology harmonization or knowledge harmonization to enable interoperability between existing systems, standards and solutions of any level of complexity without the demand for continuously adapting/revising those specifications.

The described process can be automated. The same holds for transforming the cross-domain, harmonized, consistent informational representation of the complex business system into the different ISO/IEC 10746 views for analyzing, designing, implementing and maintaining the related ICT solution.

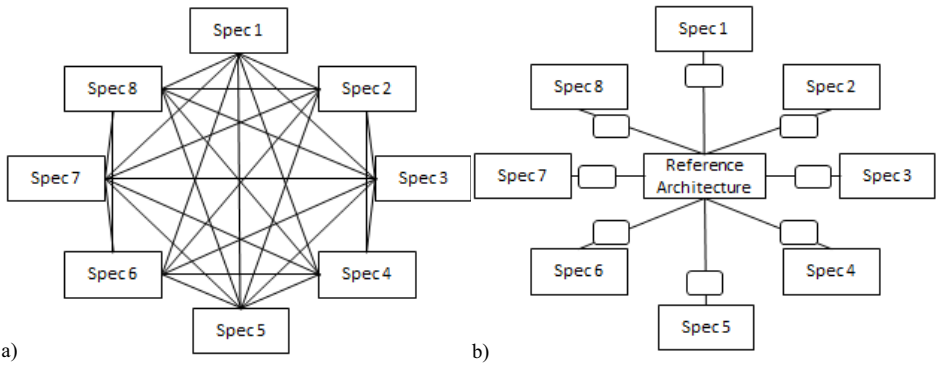


Figure 2. Standards Harmonization and Integration Approaches

Examples for re-engineering existing standards to provide cross-specification or even inter-disciplinary interoperability can be found in [5] regarding interoperability between HL7v2 and HL7v3 or in [6, 7] enabling use case and domain-crossing interoperability in the context of ISO 13972 Health informatics - Detailed clinical models. The approach has also been adopted for ISO and CEN standards such as ISO 13606-1 Health informatics – EHR communication – Reference Model, where the reference model used for all parts has been re-engineered. The feasibility of the Reference Architecture Model and Framework has also been practically demonstrated for automatically designing inter-domain Web services to facilitate multi-disciplinary approaches to Type 2 Diabetes Care management [8].

4. Discussion and Conclusions

Domain experts involved describe specific aspects of business systems in a specific context, using their specific terminologies and ontologies, methodologies and frameworks. Normally not bound to a specific framework, resulting informational representations are therefore quite inconsistent. This as well as evolving contexts or the inclusion of further domains require the adaption of existing (standardized) informational representations of domain-specific use cases, resulting in permanent revisions of specifications. Migrating to a domain-independent reference architecture developed from an abstract mathematical representation of the universe and combined with system-theoretical engineering, enables the consistent formal representation, harmonization and interrelation of any discipline to complex systems' interoperability.

The presented approach has been successfully introduced in several cross-domain ISO specifications, such as ISO 22600 Privilege management and access control, ISO 21298 Functional and structural roles, HL7 Composite Security and Privacy Domain Analysis Model. A simplification of the model is the basis of the open architectures for national health information systems in developing African countries [9]. The approach also allows a comparative analysis and evaluation of ICT Enterprise Architectures [4]. Recently, ISO TC 215 decided to include the Interoperability Reference Architecture Model in all specifications crucial for health systems interoperability such as ISO 13606, ISO 13972 or ISO 21298 Health informatics – Health systems architecture, managed by ISO TC 215 Working Group 1 "Architecture, Frameworks and Models", but also including related specification work of the other WGs.

Acknowledgement

The author is indebted to thank his colleagues from ISO TC 215, CEN TC 251, and HL7 International for their kind cooperation and support.

References

[1] Lankhorst M et al. Enterprise Architecture at Work. The Enterprise Engineering Series. Berlin, Heidelberg: Springer Verlag, 2009.
[2] Akerman A, Tyree J. Using ontology to support development of software architectures. IBM Systems Journal 2006; 45 (4): 813–825.
[3] Blobel B. Architectural approach to eHealth for enabling paradigm changes in health. Methods Inf Med. 2010; 49,2: 123-134.
[4] Blobel B, Oemig F. The Importance of Architectures for Interoperability. Stud Health Technol Inform. 2015; 211: 18-56.
[5] Oemig F. Development of an ontology-based architecture for ensuring semantic interoperability between communication standards in healthcare (in German). PhD Thesis, University of Regensburg, Medical Faculty, Regensburg, Germany, 2011.
[6] Goossen WTF. Detailed clinical models: Representing knowledge, data and semantics in healthcare information technology. Healthcare Informatics Research 2014; 20(3): 163-172.
[7] Goossen W, Gossen-Baremany A, Van Der Zel M. Detailed clinical models: A review. Healthcare Informatics Research 2010; 16(4): 201-2014..
[8] Uribe Gómez GA. An Architecture-Centric and Ontology-Based Approach to Cross-Domain Interoperability of Health Information Systems for Diabetes Care. PhD Thesis, Universidad del Cauca, Facultad de Ingenieria Electronica y Telecomunicaciones, Popayan, Colombia, 2015.
[9] Moodley D, Pillay AW, Seebregts CJ. Researching and developing open architectures for national health information systems in developing African countries. LNCS 2012; 7151: 129-139.

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-406

Querying Archetype-Based Electronic Health Records Using Hadoop and Dewey Encoding of openEHR Models

Erik SUNDVALL[a,b,1], Fang WEI-KLEINER[a], Sergio M FREIRE[c], and Patrick LAMBRIX[a,d]
[a] *Linköping University, Linköping, Sweden* [b] *Region Östergötland, Linköping, Sweden*
[c] *Universidade do Estado do Rio de Janeiro, Rio de Janeiro, Brazil*
[d] *Swedish e-Science Research Centre, Sweden*

Abstract. Archetype-based Electronic Health Record (EHR) systems using generic reference models from e.g. openEHR, ISO 13606 or CIMI should be easy to update and reconfigure with new types (or versions) of data models or entries, ideally with very limited programming or manual database tweaking. Exploratory research (e.g. epidemiology) leading to ad-hoc querying on a population-wide scale can be a challenge in such environments. This publication describes implementation and test of an archetype-aware Dewey encoding optimization that can be used to produce such systems in environments supporting relational operations, e.g. RDBMs and distributed map-reduce frameworks like Hadoop. Initial testing was done using a nine-node 2.2 GHz quad-core Hadoop cluster querying a dataset consisting of targeted extracts from 4+ million real patient EHRs, query results with sub-minute response time were obtained.

Keywords. Medical Record Systems, Computerized; Database Management Systems, Dewey encoding, Archetypes, openEHR, Hadoop, Epidemiology, XML

1. Introduction

The adoption of standardized "archetype"-based Electronic Health Record (EHR) systems is increasing globally. Such systems use a fixed *reference model* (from e.g. openEHR, ISO 13606 or CIMI) that provides basic building blocks that are then assembled and constrained (primarily by clinicians) into clinically relevant structures using modeling layers consisting of *archetypes* and *templates* [1]. This partly resembles how XML building blocks can be assembled and constrained by layers of schemas. "Archetyped" instance data, i.e. conforming to archetypes, templates and the corresponding reference model (RM), often form deep tree structures where path-based querying is useful for both single-patient and epidemiological multi-patient use cases.

Existing deep tree storage and retrieval mechanisms (for example XML- and JSON-databases) can be reused for archetype-based systems [2]. There are open source XML-database solutions fully capable of handling single-patient use-cases for databases with millions of archetype-based records [3]. However, for good performance for large population epidemiological queries, other approaches or optimizations are needed [4]:

[1] Corresponding Author: erik.sundvall@liu.se. IMT, Linköping University, 581 85 Linköping, Sweden

- If the queries are well known beforehand and regularly re-run, like for recurring statistical needs, then common solutions are a) to have database administrators (DBAs) create query-specific index optimizations or b) to export relevant parts to a data warehouse (OLAP) system.
- In exploratory epidemiological research, on the other hand, where query results often lead to new questions, the delay caused by waiting for a DBA or other constrained (often human) resources can become a major bottleneck. Thus explorations of automated optimizations also handling ad-hoc population queries have been published [4]. In this publication we describe *Dewey encoding* as another method to add to the arsenal of solutions for these ad-hoc query use cases.

Storage of archetype-based data can be implemented using various database approaches [2][3][4][5] but are often queried using the implementation-neutral AQL (Archetype Query Language) [6] that can be translated into the database's ”native” query language.
Many kinds of tree-shaped data can be *Dewey encoded* and stored in many kinds of persistence solutions. This paper contributes an outline of how Dewey indexing can be customized by including *archetype_node_id* in the index paths in order to enhance performance when applied to openEHR models and associated *AQL* queries; non-enhanced XML Dewey coding would instead store *archetype_node_id* as attribute+value.
In our implementation we used Hadoop as a scalable persistence and query execution environment. Adapting the approach for other persistence platforms and non-XML-formats should be rather straightforward.

2. Background

Based on the Dewey Decimal Classification [7], which was originally developed for general knowledge classification, Dewey order encoding is among the popular methods for encoding XML documents [8]. With Dewey order, each node in an XML tree is labeled with a vector (for example the string 1.8.3.1) representing the path from the root to the node following the rules: (1) The root is labeled as an empty string; and (2) For a non-root element *u, label(u)=label(v).x*, where *u* is the *x*th child of *v*.
With the Dewey order encoding, each node in the XML tree is labeled with a distinct string. This enables Axis operations in an XPath-style query such as ::child, ::descendant and ::sibling to be simply formulated with common string operations. Based on the Dewey order coding, all the distinct paths of the underlying XML document can then be enumerated and to each path expression, the set of the corresponding nodes (in the form of Dewey encoding) is referred. The steps above are in fact the essential operations used in the so-called XML document shredding, with the ultimate goal of storing and querying the XML data in Relational Database Systems.

3. Method

Since the collection of all archetype-based data instances in an EHR system is a giant logical tree with paths, we follow the general routine of Dewey order encoding [8] to label all the nodes in the tree. One index table (*pathindex* in Figure 1) assigns a consecutive path id number to each unique path found in the data, thus if the path has

already been seen in a previous document (EHR extract) then the path is not added to pathindex. As an example of this, consider Table 1 that shows that there were 2820 different paths in the EHR extracts of 4.2 million patients (labeled sus4200k). The Dewey id of the document node containing a path is then appended to a "TP"-table (Figure 1) that has a name based on the *Path id* number corresponding to the path in the pathindex table. If it is a branch node then the corresponding table is of type Branch, if there instead is a leaf/data/text value at the node it is of type Leaf, see Figure 1. This design using a large number of tables/files suits the distributed Hadoop implementation when data and computational load is to be distributed over several servers.

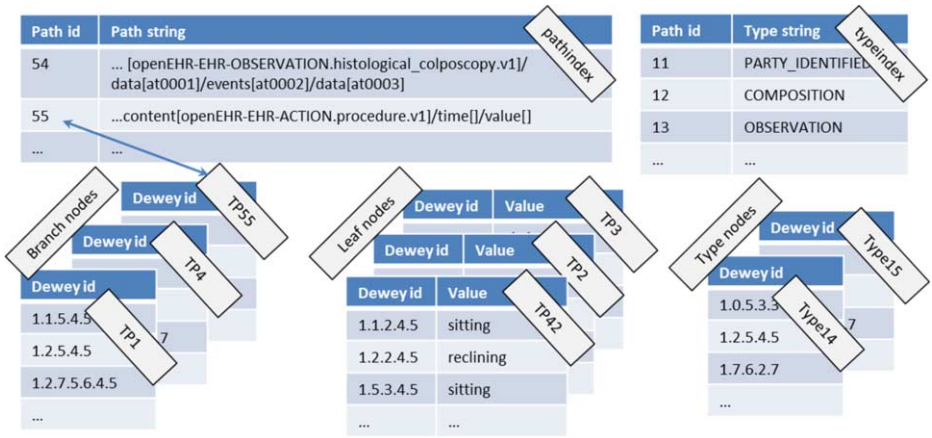


Figure 1. Simplified table examples. The number of TP-tables corresponds to the number of entries in the pathindex table that grows when new kinds of clinical information (using new archetypes etc) is added to the system. The number of rows within the TP-tables increases with the number of patients and notes per patient.

archetype_node_id: In an archetype-based tree structure, many node instances have an attribute named archetype_node_id that via archetypes identify the clinical meaning of the node [1]. In an AQL query [6], archetypes play important roles by occurring in the form of archetype predicates such as e.g. [openEHR-EHR-COMPOSITION.encounter.v1], and node predicates such as e.g. items[at0001]. Therefore, we make the archetype_node_id a first class citizen by including archetype_node_ids in the enumerated paths. An example path expression is: `...data[at0001]/events[at0002]/data[at0003]/items[at0004]/value[]...`

This is in contrast to the conventional XML transformation method [8] of storing the archetype-id and values in corresponding attributes in a table, thus a considerable amount of join operations on the attribute tables and the path tables can be avoided.

Reference Model types. RM types such as COMPOSITION, OBSERVATION occur in the FROM clause of an AQL query as class expressions to scope the data source for the query. We collect all the RM types and for each type *t*, the set of Dewey_IDs whose type is tagged *t* is stored in a table (see "Type nodes" in Figure 1).

AQL query processing. A formal description of the translation from AQL to SQL is beyond the scope of this paper. In the following we outline general rules with the help of a running query example (taken from a Brazilian epidemiology context described in [3] and [4]) which returns the records in a certain date interval where a certain examination is missing (indicated by the presence of a null-flavour value):

```
SELECT e/ehr_id/value as ehr_id FROM Ehr e
CONTAINS COMPOSITION c [openEHR-EHR-COMPOSITION.citologic_exam_form.v1]
```

```
CONTAINS OBSERVATION obs [openEHR-EHR-OBSERVATION.siscolo_anamnesis.v1]
WHERE EXISTS
obs/data[at0001]/events[at0002]/data[at0003]/items[at0004]/value
AND EXISTS
obs/data[at0001]/events[at0002]/data[at0003]/items[at0022]/null_flavour
AND (c/context/start_time/value >= $beginTime AND
c/context/start_time/value < $endTime)
```

The translation takes the following steps:

1. Identify paths. All paths can be extracted from the SELECT and WHERE clauses. Moreover, we retrieve the archetype predicates from the FROM clause and combine them together to produce the desired path expressions. For the example query we generate four paths ending with these substrings (suffixes):

P1	e/ehr_id/value
P2	[openEHR-EHR-OBSERVATION.siscolo_anamnesis.v1] /data[at0001]/events[at0002]/data[at0003]/items[at0004]/value
P3	[openEHR-EHR-OBSERVATION.siscolo_anamnesis.v1] /data[at0001]/events[at0002]/data[at0003]/items[at0022]/null_flavour
P4	[openEHR-EHR-COMPOSITION.citologic_exam_form.v1] /context/start_time/value

Note that if the retrieved path is not unique, we will retrieve a set of paths. To simplify the presentation, we assume all above paths are unique.

2. Retrieve tables. For each path from the last step, we retrieve the tables T_{P1}, T_{P2}, T_{P3} and T_{P4} respectively. Note that T_{P1} and T_{P4} contain Dewey_ID and value attributes (leaf tables), and T_{P2} and T_{P3} contain only Dewey_IDs (branch tables).

3. Filters. The filter (`value >= $beginTime AND value < $endTime`) is applied to T_{P4}.

4. Join operation. We conduct a join operation on T_{P2} and T_{P3} with the Dewey_ID of the variable `obs` as the join attribute. Note that the Dewey_ID of `obs` can be conveniently obtained by trimming a suffix of certain length on the Dewey_ID of T_{P2} and T_{P3} respectively. We name the result table as T_{P23}.

5. Containment checking. Analogously to `obs`, we can obtain the Dewey_ID of `e` and `c` from T_{P1} and T_{P4} respectively. Then the containment checking is conducted by the join operation on T_{P1}, T_{P23} and T_{P4} with the join condition that `e` is the ancestor of `c` and `c` is the ancestor of `obs`. We name the result table as T_{P1234}.

6. Type checking. We check whether `c` is of type `COMPOSITION` and `obs` is of the type `OBSERVATION` by joining T_{P1234} with the type table $T_{COMPOSITION}$ and $T_{OBSERVATION}$.

4. Results and Discussion

The data sets used for tests are sus42k, sus420k and sus4200k, which contain information about 42,428, 424,270 and 4,242,500 patients, respectively. They have been described and used for other performance tests in [4]. The data is shredded into the branch and leaf type tables. The characteristics of the datasets are given in the following table.

Table 1. Characteristics of the datasets used in this study. Size is the total size of all the shredded tables.

Shredded dataset	Size	# Paths	#Branch tables	#Leaf tables	#Type tables	Average table size
sus42k	4.5 GB	2625	1819	806	24	1.7 MB
sus420k	45 GB	2784	1929	855	24	16 MB
sus4200k	450 GB	2820	1953	867	24	158 MB

The experiments were conducted in a nine-node cluster Hadoop server; each node with quad 2.20 GHz AMD Opteron(TM) 6274 Processors running Debian GNU/Linux 6.0.10 with 12 GB RAM and 1 TB hard disk. The Hadoop block size was 64 MB. The query,

also used in [4], was manually translated and coded in Java and then executed using Hadoop map-reduce processing on the datasets described in Tables 1 and 2.

Table 2. Datasets, table sizes and results from query performance testing on a Hadoop cluster. The last column shows the running time values of the query processing. For each dataset, the query was run 10 times and an average running time value was calculated.

Table size:	T_{P1}	T_{P2}	T_{P3}	T_{P4}	Total	Running time
sus42k	2.1 MB	0.7 MB	0.1 MB	1.6 MB	4.5 MB	21 s
sus420k	22 MB	6 MB	1 MB	13 MB	42 MB	29 s
sus4200k	218 MB	60 MB	7 MB	130 MB	415 MB	59 s

The results of this initial test indicate that archetype-aware Dewey indexing can be a useful tool to support ad-hoc querying in large datasets on systems with relational capabilities (join and containment operations etc.) The indexing is related to *inverted indexes* (like Apache Lucene) that have been applied in other openEHR settings [2].

Comparisons to previous work: Performance tests [3][4][5] of different archetype-based systems are hard to compare with each other due to different data and hardware. In previous studies [3][4], using this data, Couchbase gave really fast results for recurring queries, but required indexing time for every new (unseen) query; for really small datasets BaseX was faster than this more scalable Dewey+Hadoop approach.
Future work: a) Benchmarking studies should compare several solutions on the same hardware setup using the same data and queries.
b) The single query translated manually and tested is just an initial feasibility test, but we were encouraged by other researchers to publish it so that other teams can explore it. In any realistic production system the AQL to Hadoop-querying translation should be automated instead of hand-coded in order to support the exploratory epidemiological research use case described in the introduction. Automating such translation is also a reasonable prerequisite to ease preparation of further performance tests with multiple different AQL queries. Other automated AQL translations have been done earlier [3], and AQL grammars are available [6].

References

[1] openEHR architecture overview, Release-1.0.3, 2015, http://www.openehr.org/releases/BASE/Release-1.0.3/architecture_overview.html
[2] Frade S, Freire S, Sundvall E, Patriarca-Almeida J, Cruz-Correia R. Survey of openEHR storage implementations. *26th IEEE International Symposium on Computer-Based Medical Systems* 2013 June 20–22; Porto, Portugal. p. 303–307.
[3] Freire S, Sundvall E, Karlsson D, Lambrix P. Performance of XML Databases for Epidemiological Queries in Archetype-Based EHRs. *Scandinavian Conference on Health Informatics*. 2012 October 2–3; Linköping, Sweden. Linköping Electronic Conference Proceedings 70, p. 51–57.
[4] Freire S, Teodoro D, Wei-Kleiner F, Sundvall E, Karlsson D, Lambrix P. Comparing the Performance of NoSQL Approaches for Managing Archetype-Based Electronic Health Record Data. *PLoS ONE* 11(3): e0150069, 2016.
[5] Wang L, Min L, Wang R, Lu X, Duan H. Archetype relational mapping - a practical openEHR persistence solution. *BMC Medical Informatics and Decision Making* 15:88, 2015.
[6] Archetype Query Language (AQL), http://openehr.org/releases/QUERY/latest/docs/AQL/AQL.html
[7] Dewey Decimal Classification, https://en.wikipedia.org/wiki/Dewey_Decimal_Classification
[8] Tatarinov I, Viglas SD, Beyer K, Shanmugasundaram J, Shekita E, and Zhang C. Storing and querying ordered XML using a relational database system. *ACM SIGMOD International Conference on Management of Data.* 2002 June 2-6; New York, NY, USA. ACM, p. 204-215.

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-411

HEMIC Project: Design of a Clinical Information Modelling Tool Based on ISO13972 Technical Specification

Alberto MORENO-CONDE[a,1], Francisco SANCHEZ-LAGUNA[b], Bidatzi MARIN-BASTIDA[b], Antonio ROMERO-TABARES[b], Eva MARTIN-SÁNCHEZ[b], Dipak KALRA[c] and Carlos Luis PARRA-CALDERÓN[a, b]

[a] *Group of Research and Innovation in Biomedical Informatics, Biomedical Engineering and Health Economy. Institute of Biomedicine of Seville, IBiS / Virgen del Rocío University Hospital / CSIC / University of Seville, Seville, Spain.*

[b] *Andalusian Health Service, Seville, Spain*

[c] *European Institute of Electronic Health Record, Brussels, Belgium*

Abstract. The Andalusian Health Service is the public healthcare provider for 8.302.923 inhabitants in the South Spain. This organization coordinates primary and specialized care with an IT infrastructure composed by multiple Electronic Health Record Systems. According to the large volume of healthcare professionals involved, there is a need for providing a consistent management of information through multiple locations and systems. The HEMIC project aims to address this need developing and validating a methodology based on a software tool for standardizing information contained within EHR systems. The developed tool has been designed for supporting the participation of healthcare professionals the establishment of mechanisms for information governance. This research presents the requirements and designs for of a software tool focused on the adoption of recognized best practice in clinical information modeling. The designed tool has a Service Oriented Architecture that will be able to integrate terminology servers and repositories of clinical information models as part of the modeling process. Moreover, the defined tool organizes clinicians, IT developers and terminology experts involved in the modeling process in three levels to promote their coordination in the definition, specialization and validation of clinical information models. In order to ensure the quality of the developed clinical information models, the defined tool is based on the requirements defined in the ISO13972 Technical Specification.

Keywords. Quality Management System, Tool, Software, Semantic Interoperability, Electronic Health Record.

1. Introduction

Nowadays there are multiple specifications that aim to define how clinical information is structured in order to be transferred between EHR systems. Some of the most relevant specifications are: Detailed Clinical Model, HL7 CDA templates, HL7 FHIR

[1] Alberto Moreno Conde, Grupo de Innovación Tecnológica, Hospital Universitario Virgen del Rocío, Edif. Centro de Documentación Clínica, Av. Manuel Siurot, s/n. Seville, 41013, Spain; E-mail: albertomorenoconde@gmail.com.

resources, CIMI models, ISO 13606 and openEHR archetypes. As part of this research we use the term Clinical Information Model to be able to describe any of the above presented specifications applicable to define how clinical information is structured in order to be transferred between EHR systems.

1.1. Clinical Information Modeling Processes, Quality Standards and Tools

The Clinical Information Modeling Process (CIMP) is an iterative process that includes the analysis of the domain and requirements, designing, implementing, validating and maintaining CIMs. The analysis of how the CIMs are developed in the published literature shows that it is possible to define a unified process to guide CIM definition, including the description of best practices to increase the quality of the CIMs because there are not particular differences on the process adopted associated with the applied EHR specification [1].

In the clinical information modeling field, ISO 13972 Technical Specification: Detailed Clinical Models Definition and Processes [2] describes how to implement quality processes that lead to the recommended definition of CIMs. Moreover, this specification details a set of testable quality attributes of these resulting models and how to implement a Quality Management System for the CIMP. The implementation of a Quality Management System allows establishing a continuous improvement cycle through the continuous adaptation of processes and measurements within each of the steps to obtain improved quality in the final product. Given that this specification has been recently approved, most of the existing modeling tools don't comply with the full list defined CIMP requirements.

Clinical information modeling tools are software platforms designed to support the processes associated with the definition of CIMs, as well as, establishing governance for the multiple CIMs applicable within an infrastructure or domain. Currently there are multiple tools such as the openEHR suite [3], DCM suite [4] and LinkEHR [5] that provide the mechanisms for defining and management CIMs in the form of archetypes or DCMs. A recent evaluation study shows that existing modeling tools have a good adoption of functionalities related with the management EHR specifications, data types, terminology binding and CIM metadata. As well, this study identified the need for increasing the support of the CIMP [6].

1.2. Andalusian Health Service

The Andalusian Health Service (AHS) is a public healthcare provider responsible for providing care for more than 8 million people. AHS has developed an IT infrastructure composed by multiple EHR systems covering primary and specialized care. The development of this infrastructure began in 2000, with the deployment of a centralized EHR system called Diraya in all the primary care centers of the region. Diraya contains modules that manage unique identification, ePrescription, Diagnostic orders, appointments, professional access for the region. In the subsequent years, additional EHR systems for hospital care and emergency department were developed. As a result, the AHS IT infrastructure has been deployed through the full network of 29 hospitals and 1,500 Primary care centers distributed throughout the region. The deployed infrastructure is currently used by approximately 85,000 health professionals and it contains more than 70 million ePrescriptions and 40 million primary care encounters.

The AHS has defined a central strategy for improving the quality of integrated care through the definition of a set of Integrated Healthcare Processes (IHP) for the 70 most relevant diseases. IHPs are based on the development of flexible organizational models and an appropriate management of the processes according to the integration of scientific knowledge and evaluation of their performance in healthcare environment. Each IHP describes how multiple actors are involved in each of the steps of the patient care through the multiple healthcare centres of the region.

Based on the existing infrastructure and the need for providing integrated care, AHS has strong need for establishing a consistent management of information mechanisms to ensure that information collected is adequately exploited and analyzed. The HEMIC project aims to develop and validate a methodology based on a software tool for standardizing information contained within EHR systems.

2. Methodology

This research aims to define the requirements for an online tool that support the CIMP through functionalities that ensure the application of the previously defined methodology for defining forms and information models. The definition of requirements was based on: (i) CIMP identified as part of the systematic literature review about papers talking about semantic interoperability in EHR systems [1]; (ii) metrics defined as part of the ISO13972 standard for implementing a Quality Management System [2]; and (iii) essential requirements for modeling tools [7].

3. Results

The HEMIC tool has been designed as a software instrument that will support the coordination of those healthcare professionals involved in the modeling process. This tool includes roles to classify users in three levels. The first level includes the coordinators that ensure the establishment of the information governance process. They coordinate and manage the definition of resources applicable for the multiple healthcare domains in the form generic CIMs. The second level includes a core team of multidisciplinary experts who work in depth on the detailed clinical and technical needs that the system and CIMs will need to satisfy. The third level comprises a larger group of domain experts responsible for validating the proposed clinical document or EHR form. A checklist has been designed to provide guidance about the recommended practices for the clinicians participating in the modelling process ensuring that models were based on relevant sources of information and followed appropriate validation mechanisms.

The HEMIC tool includes mechanisms for defining and managing the multiple clinical documents associated with each IHP. Moreover, associated with the evaluation of the implemented care process, the tool will support the definition of key performance indicators Most of these indicators are based on time associated with the healthcare delivery and healthcare outcomes in the patient population. Figure 1 details how multiple roles and tasks are assigned.

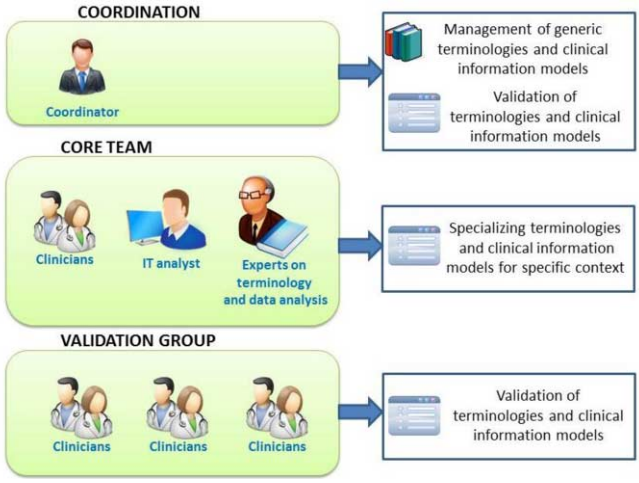


Figure 1. Coordination of participants in the definition process.

There were designed functionalities to allow clinicians designing clinical documents and the EHR forms based on the generic clinical information models approved by the coordination group. The tool specified mechanisms to access to a repository of generic CIMs and terminologies server. The Web Form Designer component accesses the list of approved semantic structures to allow clinicians define their required EHR form based on specialisation mechanisms. Specialisation mechanisms were based on the Archetype Design Principles [8] to ensure consistency of the specialised EHR form with a mechanism for detailing additional semantic context for the designed form. Associated with the definition of clinical documents, HEMIC tool generates validation tasks for modelling process coordinators ensure the satisfaction with the established information governance strategy.

The output of the modelling process will be initially in the form of CSS templates, and XML data storage structure. These outputs might be directly incorporated by the implemented AHS EHR systems as a new web form consistent with the established information governance. In addition, it will be provided an additional output according to the ISO13606 Extract structure. In the future, it is foreseen to be able to satisfy other specifications based on generic reference model such as CIMI models or openEHR. The HEMIC tool has been designed to provide the approved semantic structures to be incorporated in the EHR forms. Figure 2 shows a representation of the HEMIC tool architecture.

Figure 2. HEMIC tool architecture

4. Discussion

Traditional software development processes for defining and updating EHR systems require a large process with multiple meetings for obtaining clinical consensus and defining requirements. The traditional process has associated high cost and effort to coordinate meetings between multiple clinicians with limited availability. The HEMIC tool has been designed as an instrument focused on supporting the clinician involvement as part of the CIMP. Within the AHS there is a need for coordinating requests for modifying and generating new EHR forms, the defined functionalities are expected to allow clinicians to define clinical document and EHR forms with minimal supervision through consistent mechanisms for assigning semantics. As a consequence, clinicians would benefit from a web based tool that expects to accelerate the process of building consensus for defining the structure of EHR forms. The tool will be piloted with clinicians involved in the definition of new EHR forms as part of the AHS IT infrastructure in order to determine their acceptance and evaluate the benefits from adopting the HEMIC proposed methodology.

In order to ensure the quality of the developed CIMs, the defined CIMP has been designed to facilitate future quality labeling process based on the ISO 13972 Technical Specification.

5. Conclusion

The HEMIC tool has been designed as an instrument able to coordinate the participation of clinicians, terminology experts and IT developers as part of the CIMP based on the ISO13972 requirements. The tool includes functionalities for ensuring the establishment of information governance mechanisms through the management of generic CIMs able to be specialized to define EHR forms.

The HEMIC tool will be piloted with the definition of CIMs for one IHP in the coming months to evaluate the clinician perceived usability and acceptance. Moreover, this evaluation will assess the semantic consistency of the developed EHR forms and estimate the reduction of time in the CIMP.

References

[1] Moreno-Conde, A., Moner, D., da Cruz, W. D., Santos, M. R., Maldonado, J. A., Robles, M., & Kalra, D. Clinical information modeling processes for semantic interoperability of electronic health records: systematic review and inductive analysis. *Journal of the American Medical Informatics Association*, *22*(2015), 925-934.

[2] ISO/DTS 13972 (2015). Health informatics - Detailed clinical models, characteristics and processes, International Standardization Organization.

[3] OpenEHR Suite. http://oceaninformatics.com/solutions/knowledge_management (accessed Nov2016).

[4] DCM Suite. http://results4care.wikispaces.com/1.6+DCM_ModelCreatorENG (accessed Nov2016).

[5] LinkEHR Normalization Platform. http://www.linkehr.com (accessed Nov2016).

[6] Moreno-Conde, A., Austin, T., Moreno-Conde, J., Parra Calderón, CL., & Kalra, D Evaluation of Clinical Information Modelling Tools *Journal of the American Medical Informatics Association*, Jun 2016, ocw018.

[7] Moreno-Conde, A., Jódar-Sánchez, F., & Kalra, D. Requirements for clinical information modelling tools. *International journal of medical informatics*, *84* (2015), 524-536.

[8] Beale, T., Heard, S., (2007) Archetype Definitions and Principles. openEHR Foundation

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-416

Combining Archetypes, Ontologies and Formalization Enables Automated Computation of Quality Indicators

María del Carmen LEGAZ-GARCÍA[a,1], Kathrin DENTLER[b], Jesualdo Tomás FERNÁNDEZ-BREIS[a] and Ronald CORNET[b,c]
[a] *Facultad de Informática, Universidad de Murcia, IMIB-Arrixaca-UMU, Spain*
[b] *Department of Medical Informatics, Academic Medical Center, University of Amsterdam, The Netherlands*
[c] *Department of Biomedical Engineering, Linköping University, Sweden*

Abstract. ArchMS is a framework that represents clinical information and knowledge using ontologies in OWL, which facilitates semantic interoperability and thereby the exploitation and secondary use of clinical data. However, it does not yet support the automated assessment of quality of care. CLIF is a stepwise method to formalize quality indicators. The method has been implemented in the CLIF tool which supports its users in generating computable queries based on a patient data model which can be based on archetypes. To enable the automated computation of quality indicators using ontologies and archetypes, we tested whether ArchMS and the CLIF tool can be integrated. We successfully automated the process of generating SPARQL queries from quality indicators that have been formalized with CLIF and integrated them into ArchMS. Hence, ontologies and archetypes can be combined for the execution of formalized quality indicators.

Keywords. Clinical quality indicators, Semantic Web, Ontology, Electronic Health Record

1. Introduction

The Semantic Health Project [1] proposes the combined use of EHR standards, ontologies and terminologies for the achievement of semantically interoperable EHR systems. Current EHR standards use clinical models to describe patient data recording scenarios. The meaning of the entities in the clinical models is provided by means of ontologies and terminologies. The increasing use of ontologies for supporting EHR information and knowledge management can be illustrated through the development of frameworks. An example is our Archetype Management System (ArchMS) [2], which proposes an OWL-based semantic interoperability framework for Electronic Health Record (EHR) data, managing several repositories of archetypes, clinical data and terminologies in formats such as ADL and OWL. Archetypes are the type of clinical models used in ArchMS to share clinical data in a formal and scalable way. ArchMS

[1] Corresponding author: María del Carmen Legaz-García, Facultad de Informática, Universidad de Murcia, IMIB-Arrixaca-UMU, CP 30100, Spain; E-mail: mdclg3@um.es.

permits primary and secondary use of EHR data, e.g. for patient classification. However, it lacks specific functionality to assess the quality of care by analyzing EHR data.

Such assessment is often done using clinical quality indicators, which are "measurable elements of practice performance for which there is evidence or consensus that they can be used to assess the quality, and hence change in the quality, of care provided" [3]. Quality indicators are often released in natural language and computed manually, leading to problems of ambiguity and inefficiency. Quality indicators are often expressed as a fraction, where the denominator defines the number of patients to whom the indicator applies and the numerator defines the number of patients that received high-quality care. The CLinical Indicator Formalization (CLIF) method [4] supports users in formalizing such quality indicators in 8 steps. It has been implemented in the CLIF tool, which helps the user to create two computable queries for a given indicator, one for the numerator and one for the denominator, so that it can be computed automatically.

The computation of quality indicators often requires to integrate information from several heterogeneous sources, which is a task that can be facilitated by Semantic Web technologies. The feasibility of representing CLIF quality indicators using Semantic Web formalisms was shown in [5], but it was performed manually. In this paper we determine feasibility to automate the process by integrating the execution of CLIF quality indicators in the ArchMS framework, so the combination of archetypes, ontologies and formalization to enable the effective, automated computation of quality indicators.

2. Methods

The quality indicators created with the CLIF tool[2] are the input to our method, and SPARQL queries that can be issued over ArchMS datasets are the output. We describe next how SPARQL queries that can be executed over ArchMS datasets are generated from the 8 CLIF steps:

Step 1: encoding relevant concepts from the indicator by concepts from a terminology. The concepts have to be encoded with the ontologies used for the EHR data capture. The encoded terms are considered individuals of the class representing the clinical concept in the ontology.

Step 2: definition of the information model. This step includes the binding the concepts encoded in step 1 to entities of the information model and the definition of relations between the entities. The information model for the quality indicator is defined using an OWL ontology. For each variable included in the query a triple of the form *?variable1 a :Class1* is created, where Class1 stands for the class to which the variable belongs. The relationships between the defined variables and between the concepts defined in step 1 are specified through OWL object properties. The use of the ontology permits to assert properties.

[2] Source code available at https://github.com/kathrinrin/clif, demo at http://cliftool.org.

Step 3-6: formalization of the temporal, numeric, boolean and textual criteria. These constraints are specified in SPARQL by using filters, as described next:

- Temporal constraints are translated using FILTER, BIND and date functions such as YEAR, MONTH and DAY.
- Numerical constraints use FILTERs to compare a variable with a number.
- Textual: FILTER REGEX, using as parameter the corresponding text.
- Boolean: FILTER or FILTER NOT according to the TRUE/FALSE value.

Step 7: definition of the exclusion criteria and negations. The exclusion criteria are expressed in SPARQL using FILTER NOT EXISTS.

Step 8: identification of specific criteria for the numerator. The numerator constraints are excluded from the query of the denominator.

3. Results

3.1. Transformation Example

The application of the transformation method is illustrated next through the quality indicator "**Number of examined lymph nodes after resection of a primary colon carcinoma"**, which is defined as follows:

Numerator: number of patients with 10 or more lymph nodes examined after resection of a primary colon carcinoma.

Denominator: number of patients with lymph nodes examined after resection of a primary colon carcinoma.

Exclusion: patients with previous radiotherapy and recurrent colon carcinomas.

Step 1: encoding relevant concepts from the indicator by concepts from a terminology. The results of this step are the mappings between concepts of the indicator and the SNOMED-CT concepts shown in Table 1.

Table 1. Codification of relevant concepts of the sample quality indicator

Quality indicator concept	SNOMED CT concept
Colon resection	SCT_23968004 (Colectomy (procedure))
Primary colon carcinoma	SCT_93761005 (Primary malignant neoplasm of colon)
Lymph node examination	SCT_284427004 (Examination of lymph node)
Radiotherapy	SCT_108290001 (Radiotherapy)
Recurrent colon carcinoma	SCT_31496500 (Local recurrence of malignant tumor of colon)

Step 2: definition of the information model. The result is the OWL representation of the colon carcinoma domain ontology (see Figure 1), where the coloured classes are the ones mapped to SNOMED CT. The variables such as *?resection, ?examination or ?colonCarcinoma* are linked through the ontology:

?patient a Patient .
?resection a Procedure .
?coloncarcinoma a Diagnosis .
?coloncarcinoma hasParticipant ?patient .
?resection hasReason ?coloncarcinoma .

The codification done in step 1 is translated into the query as follows: *?coloncarcinoma isAboutSituation sct:SCT_93761005.*

Step 3-6: formalization of the temporal, numeric, boolean and textual criteria. The following temporal and numeric constraints are created:

?examination date_time_performed ?tempRel120 .
?resection date_time_performed ?tempRel220 .
FILTER (?tempRel120 >= ?tempRel220)
?resultExamination nodes_examined ?num15 . FILTER (?num15 >= 10)

Step 7: definition of the exclusion criteria and negations. The exclusion criteria for *?radiotherapy and ?recurrent* are created:

filter not exists{
?radiotherapy date_time_performed ?tempRel119 .
?examination date_time_performed ?tempRel219 .
filter(?tempRel119 <= ?tempRel219) }

Step 8: identification of specific criteria for the numerator. The numerator constraints are excluded from the query of the denominator.

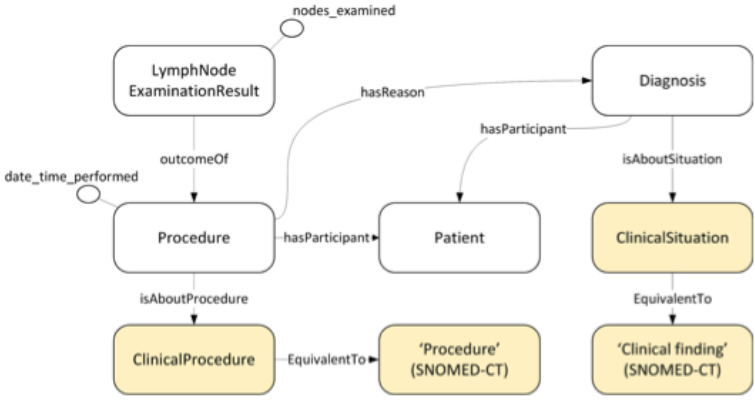


Figure 1. Excerpt of the colon carcinoma domain ontology

Figure 2. Interface of the integration of quality indicators in ArchMS

3.2. Implementation

The resulting framework is available in http://sele.inf.um.es/archms/clif. Figure 2 shows the interface to run the quality indicator queries and the result of executing our running example in a repository of 5000 patients. The resulting SPARQL queries are stored in a repository that is used by ArchMS. Given a quality indicator and a domain ontology, ArchMS recovers the SPARQL query to compute the quality indicator over its datasets.

4. Conclusions

Semantic Web technologies facilitate the interoperability of data, which helps to deal with data heterogeneity, a traditional problem of computing quality indicators. Besides, OWL reasoning may also be helpful for completing the clinical data needed for computing quality indicators. In this work, we have described the approach developed and implemented for the automated computation of CLIF quality indicators over the ArchMS semantic EHR datasets. We believe that this work constitutes an example of how the use of archetypes and ontologies can be combined to facilitate the secondary use of EHR data. In the future, we aim to compute indicators based on different EHR datasets to analyse the performance of our approach.

Acknowledgements

This work has been possible thanks to the Spanish Ministry of Science and Innovation, the Fundación Séneca and the FEDER Programme through grants TIN2014-53749-C2-2-R, and 19371/PI/14.

References

[1] D. Kalra, P. Lewalle, A. Rector, JM Rodrigues, KA Stroetmann, G. Surjan et al. (2009). Semantic interoperability for better health and safer healthcare. *Research and Deployment Roadmap for Europe. SemanticHEALTH Project Report (January 2009), Published by the European Commission, http://ec. europa. eu/information_society/ehealth*.

[2] MC. Legaz-García, C. Martínez-Costa, M. Menárguez-Tortosa, J.T. Fernández-Breis, A semantic web based framework for the interoperability and exploitation of clinical models and EHR data, *Knowledge-Based Systems* **105** (2016) 175–189

[3] M. Lawrence, F. Olesen, Indicators of Quality in Health Care, *European Journal of General Practice* **3** (1997) 103–108

[4] K. Dentler, M.E. Numans, A. ten Teije, R. Cornet, N.F. de Keizer, Formalization and computation of quality measures based on electronic medical records, *Journal of the American Medical Informatics Association* **21** (2014) 285–291.

[5] K. Dentler, A. ten Teije, R. Cornet, & N. de Keizer (2013). Semantic integration of patient data and quality indicators based on openEHR archetypes. In *Process Support and Knowledge Representation in Health Care* (pp. 85-97). Springer Berlin Heidelberg.

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-421

Parallel Design of Browsing Scheme and Data Model for Multi-Level Hierarchical Application Devoted to Management of Patient with Infectious Disease in Primary Care

Adrien UGON[a1], Catherine DUCLOS[b], Salamata KONATE[b], Sarah ARNEDOS LOPEZ[b], Hechem YAZIDI[b], Alain VENOT[b], Marie-Christine JAULENT[b], and Rosy TSOPRA[b]
[a]*Sorbonne Universités, UPMC Univ Paris 06, UMR 7606, LIP6, Paris, France*
[b]*Sorbonne Universités, UPMC Univ Paris 06, INSERM Sorbonne Paris Cité, Universite´ Paris 13, LIMICS, UMR_S 1142, Paris, France*

Abstract. Many decision systems are based on a hierarchical approach, enriching the known context used to finally choose the right potential action. Designing the scheme for browsing the clinical guidelines is a task devoted to expert in infectious diseases. Designing the data model is a task devoted to the expert in data modeling. As a consequence, browsing scheme and data model generally differ in terms of abstraction levels. While the browsing scheme proposes to navigate into depth, the data model stays flat. We propose here a novel method to design in parallel the browsing scheme and the data model so that both of them reflect the different abstraction levels in decision process.

Keywords. Clinical Decision Support System, Data model, Clinical Practice guidelines, Antibiotic Prescription, Hierarchical browsing

1. Introduction

Many Clinical Decision Support Systems (CDSSs) were developed for antibiotic prescription in primary care [6]. They implemented Clinical Practice Guidelines (CPGs), which are documents written according evidence based medicine. However the adoption of these CDSSs is often poor because of their lack of usability [5], and the time needed to get an answer [4]. We propose a new approach consisting in designing a CDSS for antibiotic prescription having an easy and consistent browsing whatever the clinical situation considered. This could allow to minimize the cognitive load of General Practitioners (GPs), and thus increase its usability and adoption.

One of the fundamental step in the conception and development of many computation based applications is the choice and the design of the data model that will

[1] Corresponding author, Adrien Ugon LIP 6 - CNRS UMR 7606, Département SoC - Equipe SYEL, B.C. 167 - 4, Place Jussieu, 75252 Paris Cedex 05, France ; E-mail: adrien.ugon@lip6.fr

be used. Even if application is based on a hierarchical approach, or a in-depth browsing, the data model stays flat. Hierarchical applications are very common, either to mine data or to abstract information [2]. Relational databases are usual models used in different applications.

The aim of the article was to define in parallel and in a similar way, a common browsing scheme for all clinical situations related to the antibiotic prescription in primary care, and a data model compliant with this browsing scheme. We will first describe the methods we followed for designing the browsing scheme and the related data model. Then we will present results obtained by applying the method to antibiotic prescription decision support system and conclude.

2. Methods

2.1. Designing the Browsing Scheme

Five CPGs related to urinary and respiratory infections were analyzed. Decision criteria used in CPGs to indicate how to manage a patient were extracted from all studied CPGs and grouped into clusters. For example, age and sex were grouped in the cluster "patient profile". Potential actions corresponding to the different possibilities of management were also then extracted and grouped into clusters. For example, "amoxicillin prescription" and "fosfomycin trometamol prescription" were grouped into the cluster "antibiotic prescription". At the end of the process, a set of decision criteria have been extracted by the domain expert and grouped into clusters with definite terminology. This terminology is devoted to be used in the data model. The browsing ends with a potential action. Given the decision criteria and potential actions previously defined, a generic and common browsing scheme, applicable for all clinical situations, was finally built.

2.2. Designing the Data Model in Accordance to Browsing Scheme

Hierarchical applications are based on different levels reflecting the different cognitive abstraction levels of the knowledge on a defined situation. Browsing entirely the application can thus be considered as assigning a value to a set of predefined variables: the decision criteria. At each level correspond a list of variables used to describe the given situation. Selecting a node to the next step consists in choosing the values for the set of variables that need to be assigned at this level in the browsing scheme. These values must then be propagated to the next page, enriching thus the context variables vector leading to a potential action. Into-depth browsing consists in selecting the right values associated to this set of decision criteria that were assigned to this level in the browsing-scheme. In practice, there is a link to go from one page to the next page and this link gathers all necessary information for the next page to be displayed properly.

For each level of the browsing scheme, a data model can be designed. This data model needs to be consistent with all decision criteria and all clusters that were previously defined by the expert at the current level. The link used to go from current page to the next page can be considered as a node, that is modeled as a transition table in the database schema. This node is thus a mapping between all situations that the expert considered for the current page. Selecting a given node has for effect to enrich the context variables sent to the next page so that it can be displayed in accordance with

the choices done by the user, and in accordance with his cognitive way to browse the application. As an example, figure 1 shows the resulting relational schema of the database following the scheme of a three-levels browsing. Basically, the transition tables is a mapping between a transition variable — that is sent to the next page — and the values of the set of decision criteria selected by the user in current and previous pages.

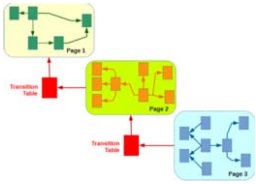


Figure 1. Example of resulting relational schema for the database compliantly with a 3-levels browsing

3. Results

3.1. Browsing Scheme Deduced from the Analysis of CPGs

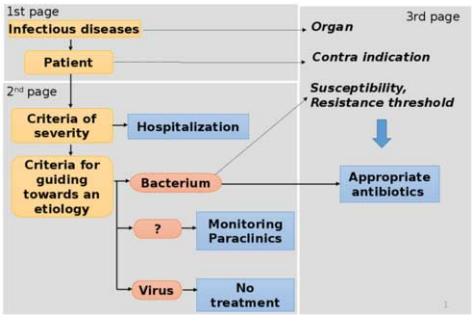


Figure 2. Browsing scheme workflow of management of patient with infectious disease in primary care

The browsing scheme workflow is showed in Figure 2. In the first page, the GP selects the infectious disease (e.g. pharyngitis), and the patient profile (e.g. adult with penicillin allergy). Then, according to the clinical situation selected, the GP can visualize the criteria of severity of the infectious disease as they are defined in CPGs. If they are present, then the patient should be hospitalized. If they are not present, then, the GP visualizes the criteria guiding towards the etiology causing the infection. Recommended treatment depends on the suspected etiology. If the suspected etiology is viral, then no treatment is needed. If the suspected etiology is unknown, then paraclinics or monitoring are recommended. If the suspected etiology is bacterial, then an antibiotic is recommended. In this last case, the GP browses through a third page displaying the appropriate antibiotics, i.e. the antibiotics active against the suspected etiologic bacterium, sufficiently concentrated in the infected organ and not contraindicated to the patient profile. This list of appropriate antibiotics is generated from criteria selected in the first and second page (e.g. patient profile in page 1).

3.2. Design of the Data Model in Accordance to the Browsing Scheme

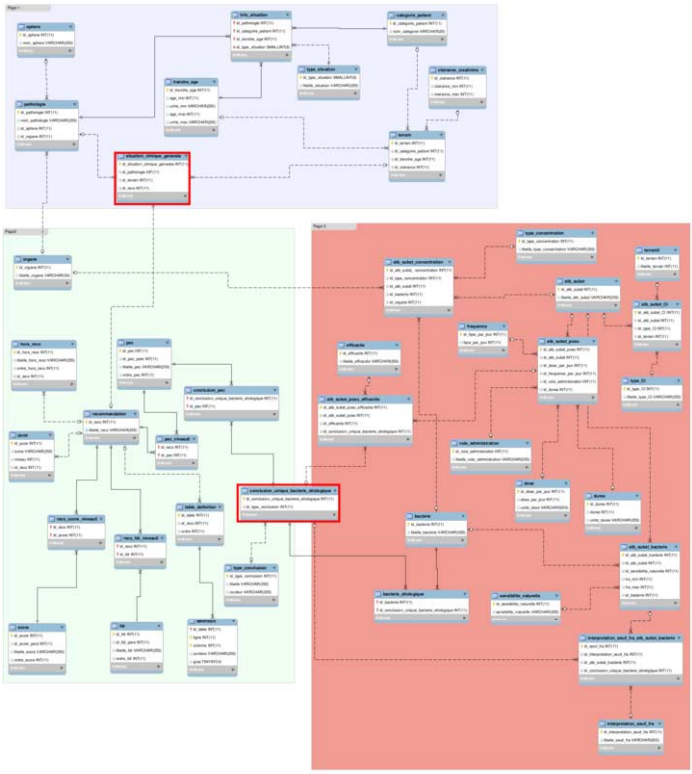


Figure 3. Database model of application for management of patient with infectious disease in primary care

To be compliant with the three-levels browsing of this web-application, we designed a relational database composed of three hierarchical levels. To make this possible, one specific data model was first designed for each level, gathering necessary information to display on the associated page. To make possible the hierarchical browsing, transition tables were added (table with red borders on figure 3), to navigate from one page to the next page by propagating previously selected information. They allow to gather all information given the current page and to send it to the next page as a contextual information. This contextual information is used to filter data to display on the page. This results into a database with relational schema - See figure 3. The database was filled in using information coming from the five CPGs used in previous step. The resulting browsing scheme was appreciated by a panel of general practitioners [7].

4. Discussion

In this work, we provided a common scheme to browse clinical guidelines and a data model reflecting the different levels used in browsing scheme. They were implemented with full respect of cognitive way for a user to choose the final potential action. Designing in the same way the browsing scheme and the data model respectful to all

intermediate steps in cognitive decision process allow to reduce the cost of design, data modeling and facilitates the development and releases of CDSSs [3,1]. Following our principles make easy the feeding and updating of the resulting database. Parallel design of browsing scheme and data model for a hierarchical application devoted to management of patient with infectious disease in primary care is an example of efficient work between medical experts and experts in data modeling. This approach needs to be investigated to confirm its applicability in other domains such as multi-level data mining or multilevel data analysis. We intend to implement the CDSS in a further work following all usability criteria defined in [7].

Acknowledgement

ANSM – Agence Nationale de Sécurité du Médicament et des produits de santé – AAP 2016 – RaMiPA

References

[1] J. S. Ash, D. F. Sittig, E. M. Campbell, K. P. Guappone, and R. H. Dykstra. Some unintended consequences of clinical decision support systems. AMIA Annu Symp Proc, pages 26–30, Oct 2007.

[2] B.V. Dasarathy. Sensor fusion potential exploitation-innovative architectures and illustrative applications. Proceedings of the IEEE, 85(1):24–38, Jan 1997.

[3] T. S. Field, P. Rochon, M. Lee, L. Gavendo, S. Subramanian, S. Hoover, J. Baril, and J. Gurwitz. Costs associated with developing and implementing a computerized clinical decision support system for medi- cation dosing for patients with renal insufficiency in the long-term care setting. Journal of the American Medical Informatics Association, 15(4):466–472, jul 2008.

[4] J. Linder, J. L. Schnipper, L. A. Volk, R. Tsurikova, M. Palchuk, M. Olsha-Yehiav, A. J. Melnikas, and B. Middleton. Clinical decision support to improve antibiotic prescribing for acute respiratory infections: results of a pilot study. AMIA Annu Symp Proc, pages 468–472, Oct 2007.

[5] A. Moxey, J. Robertson, D. Newby, I. Hains, M. Williamson, and S.-A. Pearson. Computerized clinical decision support for prescribing: provision does not guarantee uptake. Journal of the American Medical Informatics Association, 17(1):25–33, jan 2010.

[6] V. Sintchenko, E. Coiera, and G. L. Gilbert. Decision support systems for antibiotic prescribing. Curr. Opin. Infect. Dis., 21(6):573–579, Dec 2008.

[7] R. Tsopra, J.P. Jais, A. Venot, and C. Duclos. Comparison of two kinds of in- terface, based on guided navigation or usability principles, for improving the adoption of computerized decision support systems: application to the prescription of antibiotics. Journal of the American Medical Informatics Association, 21(e1):e107–e116, 2014.

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-426

Preliminary Analysis of the OBO Foundry Ontologies and Their Evolution Using OQuaRE

Manuel QUESADA-MARTÍNEZ[a], Astrid DUQUE-RAMOS[b], Miguela INIESTA-MORENO[a], Jesualdo Tomás FERNÁNDEZ-BREIS[a,1]
[a] *Universidad de Murcia, IMIB-Arrixaca-UMU, CP 30100 Murcia, Spain*
[b] *Universidad de Antioquia, Calle 70 No. 52-21, Medellín, Colombia*

Abstract. The biomedical community has now developed a significant number of ontologies. The curation of biomedical ontologies is a complex task as they evolve rapidly, being new versions regularly published. Therefore, methods to support ontology developers in analysing and tracking the evolution of their ontologies are needed. OQuaRE is an ontology evaluation framework based on quantitative metrics that permits to obtain normalised scores for different ontologies. In this work, OQuaRE has been applied to 408 versions of the eight OBO Foundry member ontologies. The OBO Foundry member ontologies are supposed to have been built by applying the OBO Foundry principles. Our results show that this set of ontologies is actually following principles such as the naming convention, and that the evolution of the OBO Foundry member ontologies is generating ontologies with higher OQuaRE quality scores.

Keywords. Ontology quality, Ontology evaluation, Ontology metrics, OQuaRE, OBO Foundry

1. Introduction

In recent years the biomedical community has increased its effort in the development of ontologies. As a consequence, the BioPortal repository[2] contains at the time of writing more than 500 biomedical ontologies. These ontologies change over time and the repositories contain all the versions. The frequency of release for new versions varies among ontologies. The availability of methods that support ontology developers in the analysis of the evolution of their ontologies would certainly contribute to improve the ontology development process, to make informed decisions about the effects of the changes made in the ontologies, and to detect whether certain modelling patterns or principles are applied. Some initiatives propose the analysis of ontologies using metrics. BioPortal calculates a set of metrics[3] for each ontology, so this provides some information about the quality of the ontology and helps ontology authors to make improvements. The analysis of ontologies based on metrics has also been used as a diagnostic task using structural, functional and usability profiling criteria [3]; using

[1] Corresponding Author: Jesualdo Tomás Fernández-Breis, email: jfernand@um.es
[2] http://bioportal.bioontology.org/
[3] https://www.bioontology.org/wiki/index.php/Ontology_Metrics

criteria such as philosophical rigor, ontological commitment, content correctness, and fitness for a purpose [5]; or presenting metrics for evaluating structural properties in the ontology [7]. All these methods are focused on the analysis of different aspects of one single ontology. In this paper we propose to apply a metrics-based approach for analysing the evolution of a set of ontologies. For this purpose, we are going to apply an adaptation of the OQuaRE framework [2] for supporting the analysis of ontology evolution processes within a common framework. The method will be applied to the OBO Foundry member ontologies. These ontologies are supposed to have been developed using the OBO Foundry principles. We are going to analyse not only the evolution of each ontology, but also the findings about the evolution of the ontologies as a group. We believe our method can contribute to generate new insights about the engineering of biomedical ontologies.

2. Materials and Methods

2.1. The OQuaRE framework

OQuaRE [1] is a framework for evaluating the quality of ontologies based on the standard ISO/IEC 25000:2005 for Software product Quality Requirements and Evaluation known as SQuaRE [4]. Briefly, OQuaRE proposes the evaluation of ontology quality using 3 levels: characteristics, subcharacteristics and metrics. The current OQuaRE version includes 8 characteristics, 29 subcharacteristics and 19 metrics. Each characteristic has a set of subcharacteristics associated that, in turn, have a set of metrics associated. The complete specification of OQuaRE can be found at[4].

Table 1. OQuaRE metrics and a brief description of how we calculate them

OQuaRE metric	Description	OQuaRE metric	Description
ANOnto	mean number of annotation properties per class	**NOCOnto**	number of the direct subclasses divided by the number of classes minus the number of leaf classes
AROnto	number of restrictions of the ontology per classes	**NOMOnto**	mean number of object and data property usages per class
CBOnto	number of direct ancestor of classes divided by the number of classes minus subclasses of Thing	**POnto**	mean number of direct ancestor per class
CBOnto2	mean number of direct ancestor per class	**PROnto**	number of subclassof relationships divided by the number of subclassof relationships and properties
CROnto	mean number of individuals per class	**RROnto**	number of usages of object and data properties divided by the number of subclassof relations and properties
DITOnto	length of the longest path from Thing to a leaf class	**RFCOnto**	number of usages of object and data properties and superclasses divided by the number of classes
INROnto	mean number of subclasses per class	**TMOnto**	mean number of classes with more than 1 direct ancestor
LCOMOnto	mean length of all the paths from leaf classes to Thing	**TMOnto2**	mean number of direct ancestor of classes with more than 1 direct ancestor
NACOnto	mean number of superclasses per leaf class	**WMCOnto**	mean length of the path from Thing to a leaf class
		WMCOnto2	mean number of path from Thing to a leaf class per leaf class

Here, we will work at the level of quality metrics (see Table 1). In OQuaRE, the values of the metrics are transformed into *quality scores* by applying scaling functions. The current version of OQuaRE uses *quality scores* in the range [1, 5]: 1 – "Not Acceptable", 2 – "Not Acceptable - Improvement Required", 3 – "Minimally Acceptable", 4 – "Acceptable" and 5 – "Exceeds Requirements".

OQuaRE offers two scaling functions [2], which differ on how the metrics values are transformed into *quality scores*, and that provide complementary information:

[4] http://miuras.inf.um.es/oquarewiki

- *Static scaling function*: based on recommendations and best practices from the Software Engineering and Ontology Engineering communities. This method uses a predefined transformation function, so the value of a certain metric is always transformed into the same *quality score*.
- *Dynamic scaling function*: based on the observed values of the quality metrics of a corpus defined by a set of ontologies. The transformation function depends on the corpus of ontologies used, so the value of a metric is transformed into *quality scores* that depends on the corpus used.

We consider that the *static scale* is more appropriate for evaluating single versions of ontologies, whereas the *dynamic scale* can provide useful information about the evolution of an ontology. OQuaRE can indeed analyse consecutive versions of the same ontology providing information about the evolution of the ontology as it is further explained in [2]. The differences between consecutive versions are captured by the *mean change*, which is calculated using the *change in scale*. Every metric shown in Table 1 suffers a *change in scale* when the *quality score* for two consecutive versions is different. A *change in scale* can therefore be negative or positive. The magnitude of the *change in scale* is the absolute value of the difference between the scores. The *mean change* accounts for those metrics with *changes in scale* and produces one value by pair of versions. In addition to this, the *accumulative mean change* provides an overview of the changes produced between non-consecutive versions. For example, if the *mean change* of four consecutive versions is 0.2, -0.4 and 0 respectively, then, the *accumulative mean change* is -0.2. Finally, the OQuaRE framework is publicly available at[5] including web services to third-parties. OQuaRE uses R for data analysis.

2.2. OBO Foundry Member Ontologies

The Open Biomedical (OBO) Foundry initiative [6] proposes the creation of an orthogonal collection of ontologies by applying shared principles[6] for the coordinated evolution of ontologies. The ontologies of the OBO repository are either member or candidate ontologies. For an ontology to be a member, the OBO Foundry must have checked that they have been developed by following such criteria. In this work we analyse the eight member ontologies (June 2015). The number of versions for some ontologies was high, so we applied a sliding window filtering algorithm using the release date. Figure 2 (columns 1-3) describes the corpus using the BioPortal acronyms.

3. Results

3.1. Comparative Analysis of the 8 OBO Foundry Member Ontologies

We describe the results of applying OQuaRE over the latest version of each ontology. The results are shown in Figure 1. There is one box by metric, the x-axis comprises the ontologies, and y-axis the *quality scores* for the *dynamic scale* (red) and for the static one (blue). Some metrics are analysed next. ANOnto is related to the "naming convention" OBO principle, which promotes the use of *rdfs:label* for the primary label

[5] http://sele.inf.um.es/oquare
[6] http://www.obofoundry.org/principles

and includes exactly one for every declared entity. Seven out of the eight ontologies get the highest score, "5", using the *static scale*. This means that more than 80% of the entities have labels. However, PR does not reach this threshold, what could be reported to its developers. The *dynamic scale* complements this result since the scale is now obtained from the actual values of the metrics of the eight ontologies. According to this scale, GO gets the highest score, followed by PATO and ZFA. The analysis of the results with the *static scale* reveals that the highest scores are obtained in metrics that represent a proper use of annotations (ANOnto), the number of elements that can be related by properties (AROnto) and hierarchical relations (CBOnto, INROnto and NACOnto). The lowest scores are for CROnto (individuals per class) and DITOnto (depth of the hierarchies) although the CROnto score can be justified by the fact that those ontologies are not supposed to have individuals. Finally, more variability between ontologies is observed for the rest of metrics.

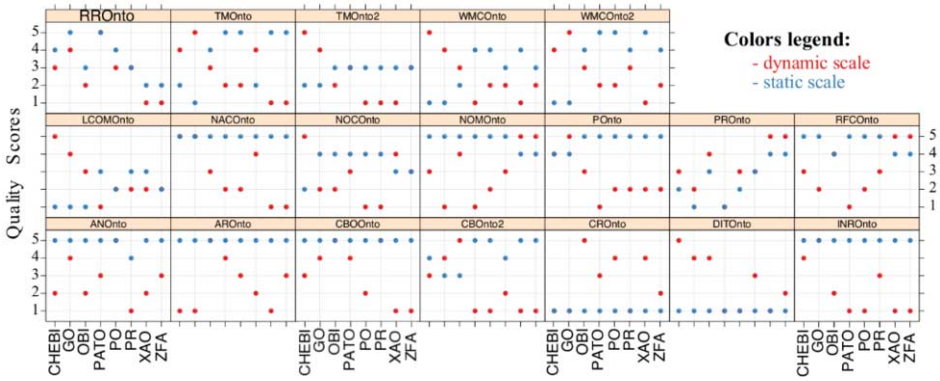


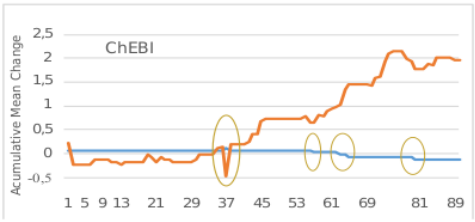
Figure 1. *Quality scores* for the latest version of each ontology in our experimental corpus.

3.2. Evolution of the Ontologies

The evolution of the ontologies has been studied by ontology. We have created eight independent corpus, including the versions of the same ontology. The trajectories of the *quality scores* for each ontology and scale can be inspected in our webpage[7]. The results of the accumulative mean change for ChEBI are shown in Figure 2. Blue and orange lines represent respectively the scores for the static and dynamic scales. ChEBI is relatively stable for the *static scale*, so this ontology does not suffer many changes in terms of *quality scores*. The four versions remarked by ovals are the ones with higher changes in the *quality scores*. If we analyse the results for the *dynamic scale*, the *accumulative mean change* is negative until version 37, turning into positive since then.

The *accumulative mean change* for the eight ontologies is shown in Figure 2 right. There, we can observe that it is negative for the *static scale*, and positive for the *dynamic scale*. Our experience reveals that the *dynamic scale* is more appropriate for analysing the evolution of ontologies, since it is able to capture smaller differences between values of the metrics. The *dynamic scale* shows that the evolution of the OBO members has produced ontologies with higher *quality scores* according to OQuaRE.

[7] http://miuras.inf.um.es/oquare/mie2017



	Versions		Accumulative Mean Change	
	Nº.	Time Period	Static	Dynamic
CHEBI	92	Jul 2008 - Jun 2015	-0.20	1.88
GO	79	Jun 2008 - Dec 2014	-0.16	2.74
OBI	17	Jul 2008 - May2015	-0.32	2.52
PATO	72	Jul 2008 - Aug 2014	-0.01	1.77
PO	9	Jan 2011 - Aug 2013	-0.05	1.57
PR	50	Mar2008 - Apr 2014	-0.67	2.24
XAO	36	May2007 - Jun 2015	-0.27	1.27
ZFA	53	Nov2005 - Mar 2015	-0.09	1.51

Figure 2. Left) Profile of the *accumulative mean change* for ChEBI ontology. Right) *Accumulative mean change* score for the different member of our corpus.

4. Conclusions

The developers of biomedical ontologies need support to analyse the quality of their ontologies and how the evolution of the ontologies is affecting to such quality. We have proposed a method based on the OQuaRE framework, which has been applied to study the evolution of the OBO Foundry member ontologies. Our results show that the ontologies follow the "naming convention" principle and that the evolution process followed is generating ontologies with higher quality scores, which is made explicit by the *accumulative mean change* metric. As future work, a deeper analysis of the relation of the OBO Foundry principles and OQuaRE metrics will be performed. We believe that this kind of method may contribute to gain insights on the engineering of biomedical ontologies and support ontology developers in generating better resources.

Acknowledgements

This work has been partially funded by to the Spanish Ministry of Economy and Competitiveness and by the Fundación Séneca through grants TIN2014-53749-C2-2-R and 19371/PI/14 respectively. Also partially funded by the FEDER Program.

References

[1] A. Duque-Ramos, J. T. Fernández-Breis, R. Stevens, and N. Aussenac-Gilles. OQuaRE: A SQuaREbased approach for evaluating the quality of ontologies. *Journal of Research and Practice in Information Technology*, 43(2):159–176, 2011.

[2] A. Duque-Ramos, M. Quesada-Martínez, M. Iniesta-Moreno, J. T. Fernández-Breis, and R. Stevens. Supporting the analysis of ontology evolution processes through the combination of static and dynamic scaling functions in OQuaRE. *Journal of Biomedical Semantics*, 7(1):63, 2016.

[3] A. Gangemi, C. Catenacci, M. Ciaramita, and J. Lehmann. Modelling ontology evaluation and validation. In *European Semantic Web Conference*, pages 140–154. Springer, 2006.

[4] ISO (International Organization for Standardization). ISO/IEC 25000:2005, Software Engineering - Software Product Quality Requirements and Evaluation (SQuaRE) - Guide to SQuaRE, 2005.

[5] J. Rogers. Quality assurance of medical ontologies. *Methods Inf. Med.*, 45(3):267–274, 2006.

[6] B. Smith, M. Ashburner, C. Rosse, J. Bard, W. Bug, W. Ceusters, L. J. Goldberg, K. Eilbeck, A. Ireland, C. J. Mungall, N. Leontis, P. Rocca-Serra, A. Ruttenberg, S.-A. Sansone, R. H. Scheuermann, N. Shah, P. L. Whetzel, and S. Lewis. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nat Biotech*, 25(11):1251–1255, 11 2007.

[7] S. Tartir and I. B. Arpinar. Ontology evaluation and ranking using OntoQA. In *ICSC '07: Proceedings of the International Conference on Semantic Computing*, pages 185–192, Washington, DC, USA, 2007. IEEE Computer Society.

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-431

Ontological Realism for the Research Domain Criteria for Mental Disorders

Werner CEUSTERS [a,b,1], Mark JENSEN[a], and Alexander D. DIEHL [a,c]
[a] *Department of Biomedical Informatics, University at Buffalo, Buffalo NY, USA*
[b] *Department of Psychiatry, University at Buffalo, Buffalo NY, USA*
[c] *Department of Neurology, University at Buffalo, Buffalo NY, USA*

Abstract. At the heart of the Research Domain Criteria for Mental Disorders is a matrix in which functional aspects of behavior are related to genotypic and (endo-)phenotypic research findings, and the various techniques through which they can been observed. The matrix is work in progress. As such it currently suffers from several shortcomings, the resolution of which, we contend, are essential to success of NIMH's goal of fostering translational science on mental disorders. Using well-established criteria for assessing the terminological and ontological quality of biomedical representations we identified the major problems to be (1) the abundant presence of terms that lack *face value*, (2) the absence of what the exact nature of the represented relationships are, and (3) referential imprecision with respect to the intended granularity of what the terms denote. We propose to eliminate these shortcomings by resorting to definitions and formal representations under the umbrella of Ontological Realism as they already have been developed in the areas of mental health, anatomy and biological functions.

Keywords. RDoC, Mental Disorders, Formal Ontology, Translational Science

1. Introduction

In 2010, the National Institute of Mental Health (NIMH) initiated the Research Domain Criteria (RDoC) project to facilitate translation of modern molecular biology, neuroscience, and behavioral approaches in an attempt to better explain the pathophysiology of mental disorders [1]. At the heart of this project is the development of a matrix in which what are called 'constructs' – some of them being further divided in 'sub-constructs' – are related to what are called 'elements', which are primarily biomarkers, such as genes and molecules, but also findings obtained through, for instance, imaging procedures or standardized questionnaires. The constructs represent functional aspects of behavior most germane to mental disorders such as the ability to receive or produce facial communication, or responsiveness to threat stimuli. They are grouped into five higher-level domains of functioning and reflect contemporary knowledge about major psychological systems: (1) negative valence systems, (2) positive valence systems, (3) cognitive systems, (4) social process systems, and (5) arousal and regulatory systems.

[1] Werner Ceusters, Departments of Biomedical Informatics and Psychiatry, Jacobs School of Medicine and Biomedical Sciences, University at Buffalo, 77 Goodell street, Suite 540, Buffalo NY 14222, USA; E-mail: ceusters@buffalo.edu.

Table 1. Biomarkers for the sub-construct 'Updating, Representation and Maintenance' of the construct 'Cognitive Control' within the Cognitive Systems domain.

Genes	Molecules	Cells	Circuits	Physiology
COMT, BDNF, DISC1 5HT2A, DRD4, DRD2 5-HTTLPR	Glu, Dopamine, GABA, NE, Acetylcholine	Pyramidal PV	DLPFC PPC Thalamocortical	Gamma synchrony; pupilometry
Behavior	**Self-Reports**		**Paradigms**	
Off-task behaviors; distractibility	Cognitive Failures Questionnaire, Disorganization Sx on SANS/SAPS/ PANSS BRIEF (Gioa)		Task Switching; AX paradigms; Cued stimulus-response reversal tasks; Tower tasks	

The incorporation of motor systems as a 6th domain is currently under debate. The elements are classified in eight groups of what are called 'units of analysis', resp. named *genes*, *molecules*, *cells*, *circuits*, *physiology*, *behaviors*, *self-reports* and *paradigms*. As an example, Table 1 gives an overview of the elements associated thus far with the sub-construct *cognitive control: updating, representation and maintenance* [2].

While the matrix is for now intended to promote the elaboration and validation of clinically relevant mental health constructs and associated measurement approaches, the hope is that it will lead to new classification schemes for mental disorders [3]. A purpose of RDoC is to stimulate research methods which avoid constraints imposed by symptom-based categorizations and synthesize interdisciplinary research in mental disorders. Therefore, the NIMH has created RDoCdb, a data repository designed to harmonize and share research and human subjects data related to RDoC and mental health [4]. Harmonization in this effort is sought by resorting to a Common Data Elements (CDE) paradigm [5]. CDEs are *metadata* constructs that have been developed to reduce time and effort spent by researchers deciding what data elements to use. They are defined in detail in a metadata dictionary so that data elements can be shared in a standardized format across multiple institutions [6]. However, although large collections of CDEs are loosely organized in contexts, they are typically not created or organized on the basis of ontological principles. Such principles require, for instance a clean separation between data and information on the one hand (brain scans, self reports, diagnoses, …), and what these data are about (brain circuits, emotions, disorders, …). They also require representations to be faithful to reality. Ignoring such principles has odd effects as exemplified in the '12-item grit scale', an RDoC approved self-report [7]. The grit-12 scores the ability of a subject to be persistent and focused in pursuit of long-term goals. Subject must rate the degree to which they self-identify with assertions, such as 'New ideas and projects sometimes distract me from previous ones'. The CDE-enabled version of the grit-12 specifies for the CDE 'interview-age': '*Age is rounded to chronological month. If the research participant is 15-days-old at time of interview, the appropriate value would be 0 months. If the participant is 16-days-old, the value would be 1 month*'.

2. Methods

The purpose of the work presented here was to assess the extent to which the matrix is congruent with terminological and ontological principles and to provide suggestions for remediation to better serve clinical and translational research in mental health. The matrix was browsed as available in [2] during October 2016. To make analysis easier,

all constructs and elements were copied into a single spreadsheet, thereby preserving their respective classification in functional domains and units of analysis as well as the definitions and explanations provided. Terms and definitions were then evaluated in function of well-known quality assessment criteria and recommendations for terminologies [8] and ontologies [9] such as face validity of terms, fixed meaning, clean separation of subsumption relations from other relations, etc. Deviations thereof were classified in coherent groups. Finally, a further literature study was conducted to identify ontological theories that could serve as candidates for adding more rigor to the matrix.

3. Results

A major problem with the current incarnation of the RDoC matrix is that for several *element* terms it is hard to assess whether they lack face value – i.e. the display term does not capture what is meant – or are erroneously classified. We found that 12 elements, e.g. dopamine and norepinephrine, appear in columns for both *genes* and *molecules* units of analysis. Clearly, nothing which is a gene can also be a molecule. The terms, when appearing under *genes*, might perhaps mean something like 'genes encoding proteins which are part of the pathway which synthesizes dopamine' or 'genes encoding receptors for dopamine'. Other examples are the presence of 'cannabinoid system', 'opioid system' and 'mouse knockout models' under *genes*, but how can a system or a model be a gene?

A second problem is that most pages accessible through [2] omit to specify what exactly are the relationships between the elements on the one hand, and the constructs and units of analysis on the other hand. Few contain references to one or more papers but without annotations from the latter to the former. More detail was provided in an earlier version of the matrix [10], what allowed us to conclude that the relationships are indeed quite diverse. For example, for the construct 'Loss', defined *as 'a state of deprivation of a motivationally significant con-specific, object, or situation*', we find in the matrix version in [2] the molecules 'glucocorticoid receptors' and 'CRH', whereas the version in [10] refers explicitly to '**down**regulation of glucocorticoid receptors' and '**up**regulation of CRH' [bold emphases added]. This is, once more, an example of the violation of the terminological principle that terms should have face value [8].

A third problem is the lack of referential precision and the overlap amongst the various units of analysis, and amongst these units and the constructs, in part caused by this imprecision. Why does the matrix refer to 'neurons' as elements under 'cells' as this seems to be too vague a cell-type in the context of mental disorders, applicable to most constructs? Yet it is only associated with 'acute threat: fear'. There is also overlap between cells and circuits, circuits and physiology, and between all of these and the constructs themselves. This is because the constructs are not well defined enough to unravel this overlap. 'Animacy perception', e.g., is a sub-construct defend as 'the ability to appropriately perceive that another entity is an agent (has a face, interacts contingently, exhibits biological motion)', while the term 'ability to appropriately attribute animacy to other agents' is used as an element belonging to the unit of analysis 'Behavior'.

4. Discussion

The RDoC initiative is a clear move away from the phenomenological "lumping" approach of the Diagnostic and Statistical Manual of Mental Disorders (DSM) and aims

to integrate a more dimensional approach anchored in neuroscience [1, 11]. It is acknowledged that it is work in progress, and that refinement is needed [3]. Our analysis indicates that there is a need for this refinement and to adopt formal ontological principles as already have been suggested in the domain of mental health [12-14].

A first step would be to reformulate the definitions of RDoC constructs and the domains under which they are classified along the lines of 'bodily systems' [15]:

X is a bodily system for organism Y if and only if: (D1)

(i) X is an element of Y;

(ii) X has a critical function for Y;

(iii) X is not a part of any other system that has a critical function for Y.

F is a critical function for system Y if and only if: (D2)

(i) some element X of Y has F as its function;

(ii) the continued functioning of system Y is causally dependent on the continued performing of F by X.

Clause (ii) in D1 offers a perspective to express formally what the functions of the systems represented by the constructs are, as well as the relationships with the various units of analysis by means of which the realization of these functions and the participation of molecules, genes, cells and circuits therein can be measured. Furthermore, the fact that bodily systems are defined in relation to critical functions, does not mean they are not related also to other functions that are not critical. Adherence to clause (iii) can reduce the observed overlap in parallel with a formal description of the parthood relationships between cells, circuits and bodily systems as represented in the FMA [16]. Caution is required, however, since D1, because of clause (iii), reserves the term 'bodily system' for the highest level systems with respect to parthood. Mapping the RDoC constructs and sub-constructs to bodily systems in the D1 sense requires thus a certain ontological commitment on behalf of the RDoC-matrix authors with respect to the precise relationships between constructs and the domain they belong to, and between sub-constructs and the constructs of which they are declared to be sub-constructs of. For example, the use of the plural in the domain 'Negative Valence Systems' suggests that the constructs 'Acute Threat ("Fear")' and 'Potential Threat ("Anxiety")' are distinct ***types*** of negative valence systems, and not ***parts*** of what would be '***the*** negative valence system' of the human body (as in 'the circulatory system'). However, how the sub-constructs 'Reward Valuation' and 'Effort Valuation / Willingness to Work' ontologically relate to the construct 'Approach Motivation', which they appear under – as parts or types, if they are systems at all and not functions – can only be determined via careful analysis by ontologists in collaboration with neuroscientists.

A second step would be to improve how various 'elements' are grouped in the matrix by defining explicitly the eight 'units of analysis', in function of what it means to be an element as defined in the context of the bodily systems and subsystems of organisms [15] (the use of the term 'element' in both [15] and the RDoC-matrix is coincidental):

X is an element of Y if and only if: (D3)

(i) X and Y are parts of an organism;

(ii) X is lower on the spatial-functional hierarchy than the organism as a whole, and lower than the system of which it is an element;

(iii) X has one or more specific functions;

(iv) X is causally relatively isolated from the parts of the organism that surround it;

(v) X is maximal, in the sense that it is not a proper part of any item on the same level of the spatial-functional hierarchy satisfying (i) to (iv).

5. Conclusion

The RDoC initiative has been received both with enthusiasm (by neuroscientists) and skepticism (by traditional psychiatrists) and several caveats have been raised from within the domain [3], primarily for how the matrix is organized and the lack of expressiveness. We demonstrated that present qualms are not ungrounded, but that improvements can be achieved by applying appropriate terminological and ontological principles grounded in Ontological Realism. Needed steps are refinement of the constructs and units of analysis, to formalize their ontological foundation, as we have exemplified by our proposal to resort to a formal definition for bodily system. Such formalization is essential for data harmonization. However, this requires a close collaboration between psychiatrists, psychologists and neuroscientists on the one hand, and ontologists on the other.

Acknowledgement

This work was supported in part by Clinical and Translational Science Award NIH 1 UL1 TR001412-01 from the National Institutes of Health.

References

[1] Cuthbert, B.N. and T.R. Insel, *Toward new approaches to psychotic disorders: the NIMH Research Domain Criteria project.* Schizophr Bull, 2010. **36**(6): p. 1061-2.
[2] NIH. *RDoC Matrix*. 2016 [cited 2016 Oct 31, 2016]; Available from: https://www.nimh.nih.gov/research-priorities/rdoc/constructs/rdoc-matrix.shtml.
[3] Kozak, M.J. and B.N. Cuthbert, *The NIMH Research Domain Criteria Initiative: Background, Issues, and Pragmatics.* Psychophysiology, 2016. **53**(3): p. 286-97.
[4] National Institute of Mental Health. *RDoCdb*. 2016 Oct 31; Available from: https://data-archive.nimh.nih.gov/rdocdb/about.
[5] National Institute of Mental Health, *Request for Information (RFI): Building a Set of Recommended Tasks and Measures for the RDoC Matrix* in *NOT-MH-16-007*. 2016.
[6] Vawdrey, D.K., et al., *Enhancing electronic health records to support clinical research.* AMIA Jt Summits Transl Sci Proc, 2014. **2014**: p. 102-8.
[7] NIMH Data Archive. *12-Item Grit Scale*. 2016 [cited 2016 Nov 3]; Available from: https://ndar.nih.gov/data_structure.html?short_name=grit01.
[8] Cimino, J.J., *Desiderata for controlled medical vocabularies in the twenty-first century.* Methods of Information in Medicine, 1998. **37**(4-5): p. 394-403.
[9] Arp, R., B. Smith, and A. Spear, *Building Ontologies with Basic Formal Ontology*. 2015, Cambridge, MA: MIT Press.
[10] National Institute of Mental Health. *RDoC Constructs*. 2016 [cited 2016 Oct 31]; Available from: https://www.nimh.nih.gov/research-priorities/rdoc/rdoc-constructs.shtml.
[11] Patrick, C.J. and G. Hajcak, *Reshaping clinical science: Introduction to the Special Issue on Psychophysiology and the NIMH Research Domain Criteria (RDoC) initiative.* Psychophysiology, 2016. **53**(3): p. 281-5.
[12] Hastings, J., et al., *The Emotion Ontology: Enabling Interdisciplinary Research in the Affective Sciences.* Modeling and Using Context, 2011. **6967**: p. 119-123.
[13] Ceusters, W. and B. Smith, *Referent Tracking for Treatment Optimisation in Schizophrenic Patients.* Journal of Web Semantics - Special issue on semantic web for the life sciences, 2006. **4**(3): p. 229-36.
[14] Ceusters, W. and B. Smith, *Foundations for a realist ontology of mental disease.* Journal of Biomedical Semantics, 2010. **1**(10): p. 1-23.
[15] Smith, B., K. Munn, and I. Papakin, *Bodily systems and the spatial-functional structure of the human body.* Stud Health Technol Inform, 2004. **102**: p. 39-63.
[16] Nichols, B.N., et al., *Neuroanatomical domain of the foundational model of anatomy ontology.* J Biomed Semantics, 2014. **5**(1): p. 1.

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-436

Bridging the Semantic Gap Between Diagnostic Histopathology and Image Analysis

Lamine TRAORE [a,b,1] , Yannick KERGOSIEN[a,c] and Daniel RACOCEANU[b,d]
[a] *Sorbonne Universités, UPMC Univ Paris 6, INSERM, Univ Paris 13, Sorbonne Paris Cité, Laboratoire d'Informatique Médicale et Ingénierie des Connaissances en eSanté (LIMICS), 15 rue de l'école de médecine, Paris, France;* [b] *Sorbonne Universités, UPMC Univ Paris 6, CNRS, INSERM, Laboratoire d'Imagerie Biomédicale (LIB), 75013, Paris, France;* [c] *Département d'Informatique Université de Cergy-Pontoise, France;* [d] *Pontifical Catholic University of Peru, San Miguel, Lima 32, Peru*

Abstract. With the wider acceptance of Whole Slide Images (WSI) in histopathology domain, automatic image analysis algorithms represent a very promising solution to support pathologist's laborious tasks during the diagnosis process, to create a quantification-based second opinion and to enhance inter-observer agreement. In this context, reference vocabularies and formalization of the associated knowledge are especially needed to annotate histopathology images with labels complying with semantic standards. In this work, we elaborate a sustainable triptych able to bridge the gap between pathologists and image analysis scientists. The proposed paradigm is structured along three components: i) extracting a relevant semantic repository from the College of American Pathologists (CAP) organ-specific Cancer Checklists and associated Protocols (CC&P); ii) identifying imaging formalized knowledge issued from effective histopathology imaging methods highlighted by recent Digital Pathology (DP) contests and iii) proposing a formal representation of the imaging concepts and functionalities issued from major biomedical imaging software (MATLAB, ITK, ImageJ). Since the first step i) has been the object of a recent publication of our team, this study focuses on the steps ii) and iii). Our hypothesis is that the management of available semantic resources concerning the histopathology imaging tasks associated with effective methods highlighted by the recent DP challenges will facilitate the integration of WSI in clinical routine and support new generation of DP protocols.

Keywords. Histopathology image analysis, semantic annotation, formal representation.

1. Introduction

In this study, we continue our semantic cognitive virtual microscopy initiative[2,3] by proposing a sustainable way to bridge the content, features, performance and usability gaps [1] [2] between histopathology and WSI analysis. The MICO project achieved a

1 Corresponding author, Lamine TRAORE, 15 rue de l'école de médecine, Paris, France; E-mail: laminet@gmail.com

2 MICO project (COgnitive MIcroscopy) - http://daniraco.free.fr/projects.htm

3 FlexMIm project (Collaborative Pathology) - http://www.systematic-paris-region.org/en/projets/flexmim

prototype system to perform some histopathology diagnosis related tasks on tissue slides where elementary imaging processes were combined by a logic engine [3], which could use formalized knowledge available as a set of rules. These rules, however, had been elaborated through local collaboration between pathologists and image scientists whereas sustainability calls for the use of publicly available knowledge gathered in standard formats from collaborative multi-centric efforts and periodic updates. A preliminarily work in this direction has been recently published by our team [3] proposing the use of the CAP organ-specific CC&P. Based on NCBO Bioportal and UMLS semantic types, the semantics generated represents a sustainable vocabulary, dedicated to histopathology, being able to effectively support daily work on WSI, in DP. Semantic models and reference terminologies are essential in DP, being able to support the reproducibility and quality of the diagnostic, to assist and standardize anatomopathological reporting, and to enable multi-center clinical collaboration or research, especially in the context of cancer grading [4].

2. Materials and Methods

The overall approach is presented in Figure 1. In this paper we treat the image analysis domain. Series of international benchmarking initiatives [5] have been launched for mitosis detection at MITOS 2012 (continued by AMIDA 2013, MITOS 2014 and TUPAC 2016), nuclear atypia grading at ATYPIA 2014 and glandular structures detection GlaS 2015. These initiatives allow envisaging a consolidated referential-database for DP.

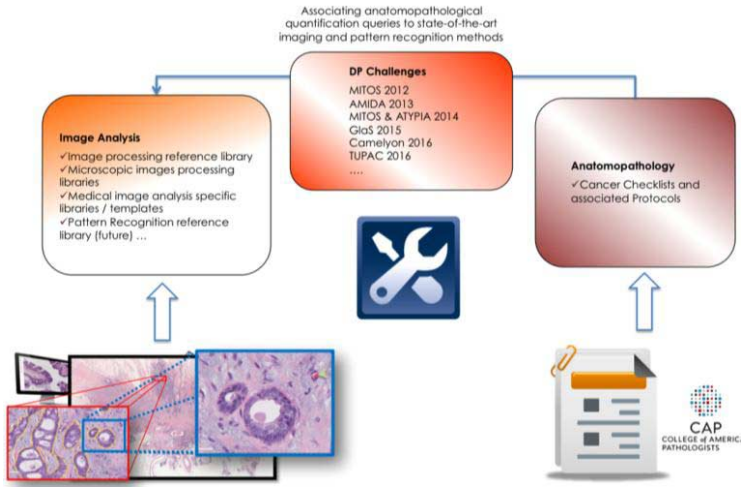


Figure 1. The overall proposed approach: use of the recent DP challenges to make an operational, instantiated link between anatomopathology and imaging.

2.1. Automatic Annotation of Corpus Issued from Contests with Available Semantic Resources in the NCBO Bioportal.

We considered the 2012-2016 period and identified 5 international benchmarking contests related to 29 top performing histopathology-imaging methods. In accordance

with our recent published work [3], 3 of the challenges are related to the breast cancer diagnosis and prognosis criteria. Corpuses were extracted from authors descriptions in articles, and "Grand Challenge" platform [5]. Table 1 summarizes description of the corpus with associated contests and methods.

Table 1. Description of the corpus with associated contests, identified methods and word count.

Corpus index	Associated conference	Identified challenges	# of methods	Word counts
C#1	ICPR 2012	MITOSIS (Mitosis detection in breast cancer histological images)	4	181
C#2	MICCAI 2013	AMIDA (Assessment of algorithms for mitosis detection in breast cancer histopathology images)	11	405
C#3	ICPR 2014	MITOS-ATYPIA (Detection of mitosis and high-grade atypia nuclei in breast cancer histology images)	4	627
C#4	MICCAI 2015	GlaS (Gland Segmentation in Colon Histology Images)	6	501
C#5	ISBI 2016	Camelyon16 (cancer metastasis detection in lymph node)	4	896

For each corpus, by using Recommender of NCBO Bioportal we obtained the ranking of the most pertinent ontologies individually or by sets of 4. The ontology-ranking algorithm used by Recommender evaluates the adequacy of each ontology to the input corpus using a combination of four evaluation criteria: Coverage, Acceptance, Detail of knowledge and Specialization. For each case, we adjusted these parameters by considering default weights (Coverage=0.55, Acceptance=0.15, Knowledge Detail=0.15, Specialization=0.15) and a focus on the coverage criterion (Coverage=1, others put to zero). We first annotated each corpus with the "imaging category" ontologies (n = 15) specified in NCBO Browse Tab. Then we redid the annotation by referring to "All ontologies" available (n = 668). In each case, the first 5 single ranked ontologies and the highest ranked ontology set (4 per set) were identified. Table 2 and Table 3 report the results.

2.2. Visual Representation of the Imaging Knowledge Issued from MATLAB, ITK and ImageJ

Our visualization targeted the concepts issued from the three image analysis communities related to the use of MATLAB (image scientists and engineers), ITK (developers) and ImageJ (imaging biologists). We used corpuses extracted from the user manuals. Conserving the hierarchy levels from sources, we organized all identified concepts with Protégé®. Then, with its OWLviz plugin we generated a visualization of concepts related to each source.

3. Results

3.1. Automatic Annotation of Corpus Issued from Contests with NCBO Bioportal Resources

3.1.1. Automatic Annotation with the 15 NCBO "Imaging Category" Ontologies

The list of five (5) most pertinent "imaging category" ontologies found in Bioportal is reported in Table 2. Overall 10 ontologies were found ranked with respect to their popularity (number of visits). From NCBO "imaging category" ontologies, the

maximum final annotation scores obtained with the coverage criterion (Coverage=1, others put to zero) were with Corpus#1: 9.0% for single ranked ontology (EDAM-BIOIMAGING) and 21.8% for ontology sets (EDAM-BIOIMAGING, NIDM-RESULTS, NEMO and IDQA). With the default configuration, single ranked ontology scores range from 11.4 (BIRNLEX) to 21.7% (NEMO).

Table 2. Top five "imaging category" ontologies found in Bioportal with associated definitions and metrics

#	Name	Category	Classes
1	Radiation Oncology Ontology (ROO)	Development, Health, Human, Imaging, Vocabularies	1183
2	DICOM Controlled Terminology (DCM)	Imaging	3476
3	Information Artifact Ontology (IAO)	Biomedical Resources, Imaging, Other	180
4	Biomedical Informatics Research Network Project Lexicon (BIRNLEX)	Anatomy, Imaging	3580
5	Neural ElectroMagnetic Ontology (NEMO)	Anatomy, Biological Process, Experimental Conditions, Human, Imaging	1851

3.1.2. Automatic Annotation with All 668 Ontologies Available on the NCBO Platform

From the results of the annotation with all ontologies available in NCBO Bioportal, we get the list of the ten (10) most relevant ontologies (with respect to their final scores) to be used for the annotation of the corpus describing imaging methods in histopathology domain. Table 3 reports the list of five (5) most relevant ontologies with related definitions and metrics.

Table 3. List of five most relevant biomedical ontologies in NCBO Bioportal for the annotation of corpus describing imaging methods in histopathology domain

#	Name	Category	Classes
1	Logical Observation Identifier Names and Codes (LOINC)	Health	187123
2	Material Rock Igneous (MATRROCKIGNEOUS)	Upper Level Ontology	3535
3	Medical Subject Headings (MESH)	Health	261990
4	Material Natural Resource (MNR)	Upper Level Ontology	3554
5	National Cancer Institute Thesaurus (NCIT)	Vocabularies	118941

Then with each of the 5 corpuses, we calculated the final score of the most relevant ontology set annotation by referring to "Imaging category" ontologies and "All ontologies" in NCBO Bioportal. By considering the same example mentioned previously with Corpus#1 and the coverage criterion final results are 57.7% for single ranked ontology (NCIT) and 75.2% for ontology sets (NCIT, SNOMEDCT, SWEET and LOINC).

3.2. Visual Representation of Concepts from MATLAB, ImageJ and ITK

Three (3) graphical tree representations reflecting the hierarchy and granularity of each source were obtained with respectively 565 concepts from MATLAB, 348 from ITK and 259 from ImageJ.

4. Discussion and Conclusion

From the above results, we report that there is no ontology related to the imaging domain in NCBO Bioportal to annotate efficiently the identified histopathology imaging methods. With respect to the ontology lists in Table 2, Table 3 and annotation results, we see that the most relevant ontologies annotating imaging concepts in Bioportal are SNOMEDCT, NCIT and other ontologies related to health, anatomy and similar categories. One should note that these huge resources are not specialized to the imaging domain. This also shows the need of an imaging domain ontology that will be built upon available image analysis concepts and functionalities.

On another hand, we faced difficulties in getting "bigger" corpus. We could find few published papers in open access, describing contest's newly proposed methods. To complete this list, we sent requests to authors to obtain more descriptions. However, in some cases patent restrictions limit the depth of the description related to the methods.

By using Protégé® and OWLviz, we obtained the visual representation of concepts issued from Matlab, ImageJ and ITK image analysis communities. This helped us to better understand the hierarchy, respective definitions and granularity of the information contained in each source. Details related to all these materials are available upon request. This work opens the perspectives of the Practical Image Processing Task Ontology (PIPTO) construction. PIPTO aims at capturing image domain knowledge in a generic way and providing a consensual understanding of concepts and functionalities identified in the standard tools from these communities.

Overall, we could identify and evaluate relevant ontologies associated to histopathology image analysis. Then by considering concepts from main biomedical imaging tools, we could propose a formal representation of the imaging knowledge from MATLAB, ImageJ, and ITK. Each of these software applications or libraries includes a set of concepts, definitions, functions and relations that are expected to cover most of the imaging methods. Additional efforts are needed to achieve a workable standard-based formal representation that will be clearly understandable by humans, machine processable, and sustainable.

5. References

[1] T. M. Deserno, S. Antani, and R. Long, "Ontology of gaps in content-based image retrieval," *J. Digit. Imaging*, vol. 22, no. 2, pp. 202–215, Apr. 2009.

[2] A. E. Tutac, *[Formal representation and reasoning for microscopic medical image-based prognosis] : [application to breast cancer grading]*. Besançon, 2010.

[3] L. Traore, C. Daniel, M.-C. Jaulent, T. Schrader, D. Racoceanu, and Y. Kergosien, "A sustainable visual representation of available histopathological digital knowledge for breast cancer grading," *Diagn. Pathol.*, vol. 2, no. 1, Jun. 2016.

[4] C. Daniel *et al.*, "Standards and specifications in pathology: image management, report management and terminology," *Stud Health Technol Inf.*, vol. 179, pp. 105–122, 2012.

[5] "Grand-challenges " [Online]. Available: https://grand-challenge.org/. [Accessed: 25-Oct-2016].

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-441

The BioTop Family of Upper Level Ontological Resources for Biomedicine

Stefan SCHULZ[a,1], Martin BOEKER[b] and Catalina MARTINEZ-COSTA[a]
[a,b] *IMI – Medical University of Graz, Austria*
[a] *IMBI – University of Freiburg, Germany*

Abstract. BioTop is a domain upper level ontology for the life sciences, based on OWL DL, introduced ten years ago. This paper provides an update of the current state of this resource, with a special focus on BioTop's top level, BioTopLite, which currently contains 55 classes, 37 object properties and 247 description logics axioms. A bridging file allows harmonising BioTopLite with the classes of Basic Formal Ontology BFO2. The updated OWL resources are available at http://purl.org/biotop. They build the core of several upper level ontological artefacts including bridging ontologies to other upper level resources.

Keywords. Upper-level Ontology, Semantic Interoperability, Description Logics

1. Introduction

Ontology engineering requires knowledge (i) in the domain to be represented, (ii) in the representation language, (iii) in the discipline of Applied Ontology [1], and (iv) in building software following design specifications. Upper-level ontologies (ULOs) guide this process by supporting the creator of domain ontologies with a consistent framework, which can guarantee for interoperability if used by different domain ontologies.

Although ULOs are often seen as domain-independent, popular ULOs like DOLCE [2] and BFO [3] target cognitive sciences and natural science, respectively. In the past, several ULOs have been created explicitly for biology and medicine, e.g. the GALEN upper level [4], the UMLS SN [5], the OBO RO [6], GFO-BIO [7], and SIO [8].

Whether ULOs positively influence the resulting artefacts or whether they contribute to excessive complexity that does not pay off from an application point of view, has been controversially discussed for two decades. However, many application ontologies exhibit a rather naïve and haphazard makeup that does not commit to any pre-existing ULO. "Quick and dirty" solutions may solve immediate problems, but they fail when it comes to maintenance, extension, modularisation and re-use. This is our rationale for ULOs. Their use requires some additional effort, involving a process towards "ontological thinking". The constraints ULOs impose on the modeller may provoke adverse reactions, but they keep them on the track and prevent them from committing modelling errors. This is the current mission of the biological upper level ontology BioTop and its overarching component BioTopLite. The objective of this paper is to provide an update of BioTop, ten years after its first publication and use.

[1] Corresponding author: Stefan Schulz, Medical University of Graz, Austria; E-mail: stefan.schulz@medunigraz.at

2. Evolution

BioTop was first launched in 2006 in OWL DL [9]. Its design has been inspired by the analysis of the GENIA ontology for cell signalling [10]. In response to a series of fundamental design problems with this resource, BioTop was proposed as an alternative, but it soon went beyond the scope of GENIA, in order to cover a broad range of categories relevant for application in all areas of life sciences. BioTop was not intended to compete with established ULOs, but rather to integrate with them. Therefore, its developers created bridging ontologies to DOLCE, BFO, and RO, and left the uppermost level deliberately flat, thus lending itself as a top-level layer for domain ontologies, without coercing developers into a certain, pre-existing ULO.

BioTop has always set a strong focus on constraining axioms, which at the time was only available for DOLCE but not for BFO and RO. This changed its scope to the integration of more classes considered fundamental for the representation of biological entities. The attempt to provide full definitions led to a further expansion into the realm of biochemistry. Experiences from the @neurIST project [11] and BioTop's use as a top-level ontology in DebugIT [12] revealed performance problems, which were mitigated by factoring out most of the chemistry classes into a separate ontology named ChemTop. This module was however not further developed due to the re-emerging ChEBI ontology [13]. Another alignment project integrated BioTop with the UMLS semantic network (SN)[14]. The resulting ontology showed, again, considerable performance problems, so that its intended use for validating UMLS sources had to be postponed. However, the task of covering the whole content of SN provided a good external criterion for content scoping [15]. BioTop's growth and ensuing performance issues motivated the creation of a “lite” version, with a minimum of content but still sufficient to link to life sciences ontologies and, nonetheless, to provide a sound framework and guidance for developers. This version was then released as BioTopLite. It was used in several experimental ontologies in which future evolutions of SNOMED CT were tested [16][17]. BioTopLite was furthermore used in experimental ontologies within the EU SemanticHealthNet project [18], in the TNM ontology [19] and in the CELDA ontology [20]. The dynamic evolution of BioTopLite, until its current version BioTopLite2, abbreviated as BTL2, led to some disruptions regarding the original BioTop ontology. Recently, however, BioTop was stripped by its upper level classes and most relations, importing BTL2, instead. Currently BTL2 has 55 classes, 37 object properties and 247 logical axioms, whereas and BioTop (without the imported BTL2 component) has 358 classes, 46 object properties and 580 logical axioms.

3. BioTop: Current Status

Figure 1 shows the current BTL2 class and relation trees. In the following, we briefly describe some of the important features:

Intuitive labels: BTL2 sets a focus on intuitive labels (*Material object*, *Quality*, *Information object*, *is part of, has participant*), which give an intuition of the meaning of the respective class or relation and is therefore better suited as "entry points" for developers compared to the class and relation names in BFO and DOLCE.

Simplified relation (object property) hierarchy: The earlier distinction between process parts and object parts, as well as between parts and proper parts has turned out to confuse rather than to help the users. We found out that one relation pair *has part / is*

part of is sufficient. Constraining axioms, necessary to control the domains and ranges of these relations were included at the class level and as general class inclusion axioms.

The class *Condition*: There is an inherent ambiguity (or "logical polysemy", according to [21] of medical terms: "Ulcer" may denote both an ulceration process as well as its result, e.g. a gastric ulcer. "Allergy" can be interpreted as an allergic disposition or as allergic manifestation.

Figure 1. BTL2 class hierarchy (left), relation hierarchy (right)

Such category distinctions (as, e.g. proposed by OGMS [22]) are not necessary in many clinical reasoning patterns, which motivated us to add the class *Condition* as a disjunction of *Material entity*, *Process*, *Disposition*, and *Function*.

The class *Situation*: Medical discourse often refers to time segments of a (biological) life, defined by the presence of a condition (in the above sense), as empirical investigations with SNOMED CT [23] have shown [24]. E.g., "gastric ulcer" would therefore refer to the life segment, called "clinical situation" or "clinical life phase", in which a gastric ulcer process unfolds, or in which a gastric ulcer structure is present. This motivated the addition of the class *Situation*, which can therefore be used as a robust bridge to what is called disorders, diseases, findings, symptoms etc. in clinical ontologies.

The classes *Universe* and *Subatomic particle*: notorious errors in OWL ontologies derive from the confusion between existential and value restrictions ("some" and "only" in Manchester syntax). In combination with transitive relations such as '**has part**', this often leads to incorrect expressions like:

'*Cell culture*' subclassOf '**has part**' only *Cell* (1)

By adding the upper level axiom:

'*Material object*' subclassOf '**has part**' some *Subatomic particle* (2)

with the latter being disjoint from other material entities a logical error occurs, which forces the modeller to revise expression (1).

Temporally qualified continuants: In 3-dimensionalist ontologies a known issue is the representation of relations between continuants, i.e. objects that exist during time, undergo temporal change and have no temporal parts like processes or time intervals. This would require three-valued relations, such as **has part** (*a, b, t*) with time as third argument. OWL-DL does not support this. The lack of temporal qualification provokes ambiguities and may lead to wrong entailments. BioTopLite 2 mitigates lack of ternary, time-dependent relations in OWL-DL by introducing time-dependent entities, subscribing to the principle of temporally qualified continuants, still under development [25]. The class '*Particular at some time*' is of no real ontological relevance but it has proven useful as a means to enforce that instances of time-dependent classes be placed in a temporal context. Class-level axioms are such that the reasoner infers that the relata must be of the type '*Particular at some time*', cf. examples in [26]. Since 2016, the following, updated ontologies are available via http://biotopontology.github.io/ which is mapped to the URL http://purl.org/biotop
The following updated ontologies are available

- BioTopLite2: http://purl.org/biotop/btl2.owl
- BioTop (importing btl2): : http://purl.org/biotop/biotop.owl
- BioTopLite2 - BFO - bridge: http://purl.org/biotop/btl2-bfo.owl

An update of the bridge files to RO, DOLCE, and UMLS SN is under way.

4. Conclusion

Ten years' experience with BioTop has shown the need for adaptation of an upper level ontology to the user's context thus providing additional, domain-specific content. A recent step towards modularization was the re-harmonisation of the original BioTop

ontology with the "lite" version BTL2, which had more dynamically evolved, primarily by disjunctive classes, simplified relations and the definition of all entities as '*Particulars at some time*'. This approach allows differentiating, by simple means, between relations that hold at some times and those that (generically) hold at all times. A gradual update of existing bridging files to other ontologies is under way, targeting the release of a family of upper-level ontological resources for biomedicine to be released in 2017.

References

[1] Smith B. Applied Ontology: A new discipline is born. Philosophy Today 12 (29):5-6 (1998)
[2] Borgo S, Masolo C. Ontological foundations of DOLCE. In Staab S, Studer R (eds.), Handbook on Ontologies (Second Edition), Springer Verlag, 2009: 361-382.
[3] Grenon P, Smith B. (2004) "SNAP and SPAN: Towards dynamic spatial ontology", Spatial Cognition and Computation, 4: 1, 69-103.
[4] Rector A, Rogers J. Patterns, properties and minimizing commitment: Reconstruction of the GALEN Upper Ontology in OWL. EKAW*04 Workshop on Core Ontologies. http://ceur-ws.org/Vol-118/
[5] McCray AT. An upper-level ontology for the biomedical domain. Comp F Genomics. 2003; 4 (1): 80-84
[6] Smith B wt al.Relations in bio-medical ontologies. Genome Biology 2005; 6 (5): R46.
[7] Hoehndorf R, Loebe F, Poli R, Herre H, Kelso. GFO-Bio: A biological core ontology. Applied Ontology, 2008, 3 (4), 219-227.
[8] Dumontier M (2013). The Semanticscience Integrated Ontology (SIO) http://code.google.com/p/semanticscience/wiki/SIO .
[9] Beisswanger E, Schulz S, Stenzhorn H, Hahn U: BioTop: An upper domain ontology for the life sciences. Applied Ontology, 2008; 3 (4): 205-212.
[10] Rak, R., Kurgan, L., & Reformat, M. (2007). xGENIA: A comprehensive OWL ontology based on the GENIA corpus. *Bioinformation*, *1*(9), 360.
[11] Boeker M , Stenzhorn H, Kumpf K, Bijlenga P, Schulz S, Hanser S. The @neurIST ontology of intracranial aneurysms: Providing terminological services for an integrated IT infrastructure. AMIA Annu Symp Proc. 2007;2007:56–60.
[12] Schober D et al. The DebugIT core ontology: semantic integration of antibiotics resistance patterns. Stud Health Technol Inform. 2010;160(Pt 2):1060-1064.
[13] Hastings J et al. The ChEBI reference database and ontology for biologically relevant chemistry: enhancements for 2013. Nucleic Acids Res. 2013 Jan;41
[14] Schulz S et al. Alignment of the UMLS semantic network with BioTop: methodology and assessment. Bioinformatics. 2009 Jun 15; 25 (12): i69-76.
[15] Schulz S et al. Scalable representations of diseases in biomedical ontologies. *Journal of Biomedical Semantics*. 2011, *2*(2), 1.
[16] Schulz S, Martínez-Costa C. Harmonizing SNOMED CT with BioTopLite: An exercise in principled ontology alignment. *Studies in health technology and informatics 2015*, *216*, 832.
[17] Cheetham E et al. Formal representation of disorder associations in SNOMED CT. Proc. of International Conference on Biomedical Ontology (ICBO). 2015
[18] SemanticHealthNet Network of Excellence (2013). http://www.semantichealthnet.eu/
[19] Boeker M et al. TNM-O an ontology for the Tumor-Node-Metastasis classification of malignant tumors: a study on rectal cancer.
[20] Seltmann S et al. CELDA – an ontology for the comprehensive representation of cells in complex systems. BMC Bioinformatics. 2013 Jul 17;14:228
[21] Pustejovsky J. The Generative Lexicon, MIT Press, 1995., xi+ 298pp. *Studies in English literature*, *1999*, 109-116.
[22] Ceusters W, Smith B. Foundations for a realist ontology of mental disease. J Biomed Semantics. 2010 Dec 9;1(1):10
[23] SNOMED CT. International Health Terminology Standards Development Organisation (IHTSDO). http://www.ihtsdo.org/snomed-ct
[24] Schulz S, Rector A, Rodrigues JM, Spackman K. Competing interpretations of disorder codes in SNOMED CT and ICD. AMIA Annu Symp Proc. 2012; 2012: 819-827.
[25] Grewe N et al. Permanent generic relatedness and silent change. Competition workshop at FOIS 2016.
[26] Schulz S, Boeker, M. BioTopLite: An Upper Level Ontology for the Life Sciences. Evolution, Design and Application. In *GI-Jahrestagung* 2013 (pp. 1889-1899).

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-446

Building SNOMED CT Post-Coordinated Expressions from Annotation Groups

Jose Antonio MIÑARRO-GIMÉNEZ[a,1], Catalina MARTÍNEZ-COSTA[a], Pablo LÓPEZ-GARCÍA[a] and Stefan SCHULZ[a]
[a] *Institute for Medical Informatics, Statistics and Documentation. Medical University of Graz*

Abstract. SNOMED CT supports post-coordination, a technique to combine clinical concepts to ontologically define more complex concepts. This technique follows the validity restrictions defined in the SNOMED CT Concept Model. Pre-coordinated expressions are compositional expressions already in SNOMED CT, whereas post-coordinated expressions extend its content. In this project we aim to evaluate the suitability of existing pre-coordinated expressions to provide the patterns for composing typical clinical information based on a defined list of sets of interrelated SNOMED CT concepts. The method produces a 9.3% precision and a 95.9% recall. As a consequence, further investigations are needed to develop heuristics for the selection of the most meaningful matched patterns to improve the precision.

Keywords. SNOMED CT, Post-coordination

1. Introduction

Encoding free-text clinical information with standard medical terminologies and ontologies is essential to improve many areas of healthcare: from semantic retrieval, to real-time decision support, cross-border data interoperability, and retrospective reporting for research and management [1]. SNOMED CT [2] is the largest medical terminology for coding clinical information, based on an ontological model of meaning. Its representational units (concepts) can be atomic ones such as *Bleeding*, *Stomach,* etc.) or pre-coordinated ones, which define complex expressions such as *Acute gastrointestinal hemorrhage* using logical axioms.

Both, pre- and post-coordinated expressions are based on the SNOMED CT Concept Model, although the meaning of some concepts is not always fully formalized (e.g. "severe" within *Severe pain*).

When annotating a given piece of information, e.g. a clinical text with SNOMED CT concepts by human annotators or NLP systems, there are several valid ways to encode the same meaning, depending on whether the user or systems tends to use pre-coordinated or primitive concepts. For example, the text "acute hemorrhage of the gastrointestinal tract" one annotator could have selected the primitive concepts C_1: *Acute hemorrhage* and C_2*: Gastrointestinal tract structure,* while another one could

[1] Corresponding author, Jose A. Miñarro-Giménez, Institute for Medical Informatics, Statistics and Documentation, Medical University of Graz, Auenbruggerplatz 2, 8036 Graz, Austria; Email: jose.minarro-gimenez@medunigraz.at

have chosen the pre-coordinated concept C_3: *Acute gastrointestinal hemorrhage.* Although there is a full logical definition of C_3, referring to both C_1 and C_2, it is not possible to infer C_3 by post-coordinating C_1 and C_2, since this would require SNOMED CT relations for connecting body structure and morphologic abnormality concepts with a clinical finding concept, in this case the relations **AssociatedMorphology** and **FindingSite**, respectively. SNOMED CT relations, however, are missing in the annotations. Our work aims at exploring the possibility of inferring meaningful post-coordinated expressions out of set of SNOMED CT concepts. This would require guessing the missing relations and putting them in the right place. The question is whether a relatively shallow text processing is enough for inferring the correct post-coordinated expressions. The SNOMED CT text annotations were produced within the ASSESS CT project. They resulted from annotating a collection of medical texts by domain experts [3]. The experts had to identify cohesive text chunks and assign a set of SNOMED CT concepts to them. The results are based on a previous study that demonstrated that most pre-coordinated expressions within SNOMED CT share a limited number of structural patterns [4]. Here we extend these patterns with the SNOMED CT Concept Model.

2. Methods

2.1. SNOMED CT Pattern Extraction Based on the Concept Model

The list of patterns from [4] is extended taking into account information from the SNOMED CT Concept Model. Figure 1 describes the extraction of one of the patterns based on a pre-coordinated SNOMED CT concept. A pattern is described in terms of the SNOMED CT top-level categories (*Clinical Finding*, *Body Structure*, *Procedure, Qualifier Value*, etc.). The selection of the particular top-level categories is indicated by the SNOMED CT Concept Model.

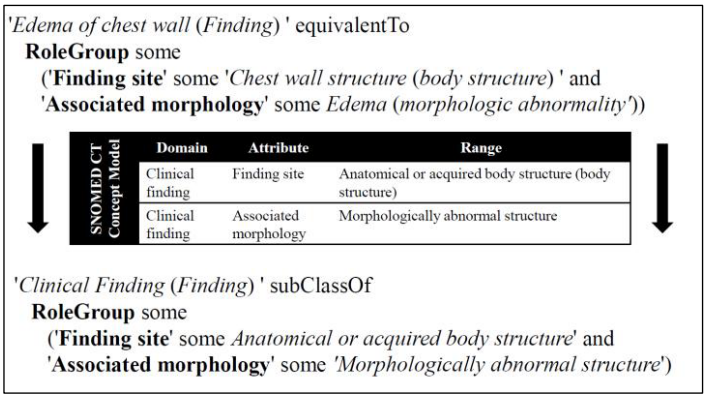


Figure 1. Post-coordination pattern extraction method uses a pre-coordinated expression (here the concept definition of *Edema of chest wall*) and the constraints from the SNOMED CT Concept Model to obtain a pattern.

2.2. Generation of Post-Coordinated Expressions

The post-coordination method takes as input the list of patterns extracted from SNOMED CT and automatically produces as output a list of valid post-coordinated expressions (not necessarily meaningful) for a provided list of AGs.

An annotation group (AG) is an unordered set of SNOMED CT concepts that jointly represent or approximate the meaning of a cohesive piece of clinical discourse, supposed to be expressible as a SNOMED CT compositional expression. Our manual annotations yielded 169 AGs. For example, the text "Given the rapid extension of the subgaleal bleeding, coagulopathy workup was initiated" was annotated with the following AG {*Subgaleal area*, *Bleeding* and *Blood coagulation panel*}.

The post-coordination method is divided in four main steps (see Figure 2). The *AG filtering* step avoids processing AGs with less than two concepts because the minimal requirement is one focus concept (i.e. the root concept within a post-coordinated expression) and one modifying concept related to it. The *pattern selection* step goes through the list of patterns to gather those whose focus concept is compatible with the focus concept in the AG. The *pattern matching* step obtains all concept combinations within an AG to match the selected patterns. Here, wherever there are two partial matched patterns with the same content, the lower frequency is discarded due to redundancy. Patterns for which all relations in the expression have a valid target concept from the AG are always retrieved. The *sorting matched patterns* step arranges the list of returned patterns based on their frequency. As a consequence, the matched patterns with higher frequencies are placed first, regardless of the matched relations and target concepts.

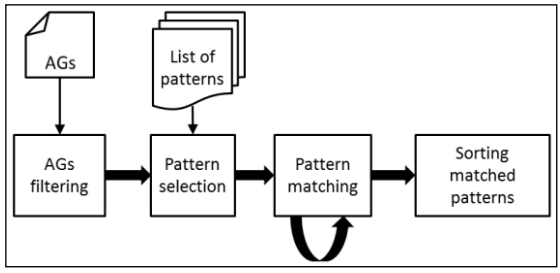


Figure 2. Post-coordination method for building SNOMED CT expression from Annotation Groups (AGs)

2.3. Evaluation Method

In order to assess whether the resulting list of patterns can be used to find meaningful post-coordinated expressions a gold standard was produced. It was manually created by a SNOMED CT expert. The post-coordination expressions of the gold standard were created based on each AG and by looking at the clinical narrative in order to assert the correct relations consistent with the SNOMED CT Concept Model. The evaluation consists in manually comparing the list of resulting matched patterns ordered by frequency with the post-coordinated expression from the gold standard. In cases with overlapped patterns, the one with highest frequency is taken. The following five categories are defined for the evaluation (with the respective counts in brackets): HIT, PARTIAL HIT, NO HIT, EMPTY, NO EXP and NO EXP AND EMPTY.

HIT means that all elements of an AG are matched against a meaningful pattern. For example, the AG {Subgaleal area, Bleeding, Blood coagulation panel} corresponds to the gold standard post-coordinated expressions:

'*Bleeding* (*finding*)' and **RoleGroup** some
('**Finding site**' *some* '*Subgaleal area* (*body structure*)')

plus the concept '*Blood coagulation panel* (*procedure*) ' which is not part of any post-coordinated expression.

NO HIT means that although some matched patterns are retrieved, none of them agrees with the gold standard. For example, the AG {*Screening mammography*, *Interval*, *Month*}, has the gold standard expression:

'*Screening mammography* (*procedure*)' and
'**Time aspect**' some *Interval* (*qualifier value*)

but if the SNOMED CT relation '**Time aspect**' does not appear in any of the patterns, therefore, no post-coordinated expression can be created.

PARTIAL HIT means that the content of the gold standard expressions are partially matched. EMPTY occurs when no patterns are matched but the gold standard provides post-coordination expressions for the AG under scrutiny.

NO EXP means that it is not possible to post-coordinate the AG elements according to the gold standard. For example, the AG {*Bisoprolol* (*substance*), *milligram (qualifier value), Twice a day (qualifier value)*}, cannot be post-coordinated because SNOMED CT Concept Model does not cover it.

Finally, we calculate the precision and recall based on the resulting matched patterns. The precision corresponds to the number of meaningful matched patterns divided by the total number of retrieved matched patterns. The recall is obtained dividing the number of meaningful patterns by the total number of expressions in the gold standard. We assume NO EXP and EMPTY are true negatives cases.

3. Results

Based on the SNOMED CT distribution files from January 2016, we obtained 956 structural patterns out of 357,165 pre-coordinated concepts with frequencies from 27,413 to 1. In particular, 284 patterns had frequency of 1. The most frequent pattern is the one depicted in Fig. 1, defining a clinical finding or disorder in terms of morphology and site.

The post-coordination method processed the list of 169 AGs in order to produce their corresponding list of matched patterns (not including AGs with only one element). It retrieved matched patterns to only 56 AGs. The AG with the most matched patterns contains nine concepts and it was associated with 377 matched patterns.

The evaluation provided the following counts (in brackets) for each category: HIT (39), PARTIAL HIT (0), NO HIT (1), EMPTY (2) and NO EXP (16), NO EXP AND EMPTY (111). As a result, we obtained a precision 9.31% and a recall of 95.9%. Moreover, the average position of a meaningful matched pattern is 7.1 in the returned list of matched pattern sorted by frequent with an average size of 30.3 patterns of list of matched patterns.

4. Discussion and Conclusion

The result from the post-coordination method shows a high rate of meaningful post-coordination expressions based on the list of patterns extracted from SNOMED CT. Consequently this list provides the most frequent patterns used for representing clinical discourse with SNOMED CT post-coordinated expressions. However, precision was very low, mainly, because the first version of the post-coordination method presented here retrieves all possible matching patterns. We have observed that the average position of a meaningful matched pattern in the resulting list of matched patterns sorted by pattern frequency is lower than the middle of the average size of the resulting list of matched patterns. The use of this frequency to limit the list of matching patterns and improve the precision will be explored in subsequent work. Yet, our method misses some meaningful post-coordination expressions, mainly for two reasons:

Firstly, there are no SNOMED CT pre-coordinated concepts that include the required pattern like in:

'*Procedure* (*procedure*)' and '**Time aspect**' some '*Time frame* (*qualifier value*)'

Secondly, the post-coordinated expression has as focus concept from the Situation with explicit context hierarchy, which has not been used within any AG, due to coding rules that guided the manual annotation task in ASSESS CT. Besides, there are several SNOMED CT relations with equivalent domain and range restrictions such as **Procedure site - Direct** and **Procedure site - Indirect**. Here, the selection of the meaningful pattern cannot be decided based on a set of concepts only, it requires analysing the semantic content of the clinical text.

Future work should allow matching patterns without focus concepts, using the top-level as default, e.g. Procedure. The focus node might then be found in previous AGs, because anaphora or ellipsis are frequent phenomena in medical texts. A necessary extension is also the addition of new structural patterns that are frequent in medical texts although they are not used within SNOMED CT definitions. This would however require a sufficiently large training set of clinical text "translated" into compositional SNOMED CT expressions, which requires time and in-depth knowledge of the SNOMED CT Concept Model.

Acknowledgements

The text annotations were performed in the EU project ASSESS CT (www.assess-ct.eu).

References

[1] O. Bodenreider. Biomedical Ontologies in Action: Role in Knowledge Management, Data Integration and Decision Support. *Yearb Med Inform*. (2008), 67–79.

[2] IHTSDO, SNOMED CT. The Global Language of Healthcare. Available: http://www.ihtsdo.org/snomed-ct [Accessed: 31-Oct-2016].

[3] J.A. Miñarro-Giménez, C. Martínez-Costa, S. Schulz. Qualitative assessment of annotations using SNOMED CT. *Proc. of the 7th Workshop on Ontologies and Data in Life Sciences* **1692** (2016), K.1–2.

[4] P. López-García, S. Schulz. Structural Patterns under X-Rays: Is SNOMED CT Growing Straight? PLOS One 2016 Nov 3;**11**(11):e0165619

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-451

HL7 FHIR: Ontological Reinterpretation of Medication Resources

Catalina MARTINEZ-COSTA[1] and Stefan SCHULZ
Institute for Medical Informatics, Statistics and Documentation, Medical University of Graz, Austria

Abstract. "A solid ontology-based analysis with a rigorous formal mapping for correctness" is one of the ten reasons why the HL7 standard *Fast Healthcare Interoperability Resources* (FHIR) is advertised to be better than other standards for EHR interoperability. In this paper, we aim at contributing to this formal analysis by proposing an RDF representation of a subset of FHIR resources based on a highly constrained top-level ontology and guided by the use of a set of Content Ontology Design Patterns (Content ODPs) for representing clinical information. We exemplify this by reinterpreting FHIR medication resources. Although a manual task now, we foresee a possible automatic translation by using RDF shapes.

Keywords. FHIR, RDF, Ontology, Semantic Interoperability, EHR

1. Introduction

The new HL7 standard *Fast Healthcare Interoperability Resources* (FHIR) [1] is a recent approach to semantic interoperability of electronic health records (EHRs). FHIR is propagated as an open standard with a high alignment with the Semantic Web [2], representing a new EHR modelling paradigm based on interoperable building blocks named *resources* [3]. FIHR resources are small data models that define a set of properties describing certain domain aspects. Currently, there are around a hundred of them, classified into six categories, and uniquely identified with a URI. Examples are *Patient*, *Practitioner*, *Medication order,* or *Observation.* FHIR resources can be serialized in JSON, XML and recently in RDF, still as a draft representation. Although FHIR was not designed with Semantic Web and RDF in mind, FHIR resources and links between them align well. The Yosemite project [5] has recently proposed RDF as universal language for healthcare data exchange. In this line, HL7 in collaboration with W3C [5] proposed an RDF representation for FHIR aiming at improving interoperability with other standards. Due to the nature of RDF and the structure of FHIR, its RDF representation focuses on representing the structure of a resource rather than the content [6]. FHIR RDF instances conform to the FHIR ontology, which introduces classes and properties. However, it is not yet connected to any formal top-level ontology such as BFO [7], BTL2 [8] or OGMS [9], and therefore ontologically shallow [10].

[1] Corresponding author: catalina.martinez@medunigraz.at.

Top-level ontologies provide domain-independent categories, relations and axioms (e.g. categories like *Process*, *Material entity*, *Quality*, etc.) in order to standardize the ontology creation by heavily constraining it, according to a rigorous ontological commitment. Therefore, building the FHIR ontology under a top-level ontology should contribute to improve its semantic interoperability with other representations.

Here we suggest the use of the top-level ontology BioTopLite 2 (BTL2) as a compromise between degree of formalization (i.e. constraining axioms) and complexity, contributing to a quick learning curve. BTL2 bridges with other ontologies such as BFO, and SNOMED CT content has started to be harmonized with basic top-level classes and relations of BTL2 [11]. Despite the benefits of top-level ontologies [11], their use is not trivial and requires of some effort. The EU project SemanticHealthNet [12] proposed content ontology design patterns (Content ODPs) to ease the modelling of clinical information under BTL2 [13]. Content ODPs provide templates for recurrent modelling content, underpinned by formal ontologies [14].

In the following we apply a set of Content ODPs to reinterpret FHIR medication resources using BTL2 and comment on the benefits of the proposed representation with a query exemplar. Finally, we discuss open issues and future work.

2. Methods

The main **FHIR Medication Resources** are shown in Figure 1 and Table 1.

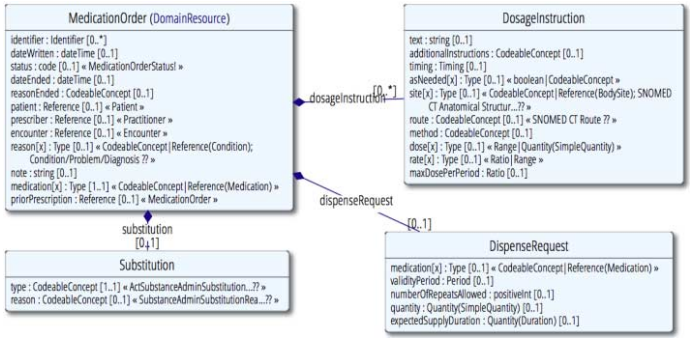


Figure 1. UML diagram of the *MedicationOrder* resource, to which the resources *DosageInstruction*, *DispenseRequest* and *Substitution* are linked. Values with "Reference" also represent FHIR resources

Table 1. Medication related main FHIR resources

Resource name	Resource description
MedicationOrder	An order for both supply of the medication and the instructions for administration of the medicine to a patient
MedicationDispense	Provision of a supply of a medication with the intention that it is subsequently consumed by a patient (usually in response to a prescription).
MedicationAdministration	When a patient actually consumes a medicine, or it is otherwise administered to them

BTL2-based Ontology Framework and Content ODPs. The integration of heterogeneous clinical information is enabled by a formal ontology framework that focuses on representing content instead of content structure, and which supports formal inference [15]. This framework encompasses the ontologies BTL2 (prefix "btl2") and SNOMED CT (prefix "sct"), as common reference point for representing the clinical content. The framework strictly distinguishes between real world content, represented

by classes like *Lung cancer*, *Blood pressure*, etc. and information (e.g. *Lab test result*, *Diagnostic statement*, *Drug order* etc.). Information entities are related to real-world entities via the relation **represents**. Content patterns act as templates to represent recurring modelling cases (e.g. Participation, Plans). Table 2 shows the main patterns for reinterpreting FHIR medication-related resources in RDF. Correspondences between OWL and RDF representations are shown in Table 4.

(1) *Plan* ISRESULTOFPROCESS *?Process*
(2) *Plan* HASINFORMATIONPART *?InformationObject*
(3) *Plan* PERFORMSPROCESS *?Process*
---
(4) *Process* HASTEMPORALVALUE *?TemporalRegion*
(5) *Process* HASRESULT *?InformationObject*
(6) *Process* HASPARTICIPANT *(?MaterialObject* or *?InformationObject*)
(7) *Process* HASQUALITY *?ValueRegion*
(8) *Process* ISDUETO (*?MaterialObject* or *?InformationObject* or *?Process* or *?ImmaterialObject* or *?Disposition*)
(9) *Process* HAPPENSBEFORE *?Process*
(10) *Process* HAPPENSAFTER *?Process*
(11) *Process* ISINCLUDEDIN *(?MaterialObject* or *?ImmaterialObject*)

Figure 2. RDF triple [16] representation of Planned Process and Clinical Process patterns. RDF predicates (in small caps) correspond to OWL object properties or expressions using BTL2. BTL2 classes are given in Italics. A question mark in an OWL class label represents a variable part within the pattern.

Table 2. Description of main Content ODPs used for describing FHIR medication resources

Content ODP name	Pattern description
Planned Clinical Process	Record entry about the intent to perform some healthcare related process (e.g. request to administer some drug, plan to reach some target body measurement (e.g. weight), request to perform some healthcare service (e.g. check potassium level, etc.)
Clinical Process	Clinical process description (e.g. observation, assessment, history taking, request process, physical examination, etc.)

3. Results

Correspondences between the medication-related resources and the RDF Content ODPs have been manually defined in order to re-interpret the existing FHIR representation based on the proposed ontology framework. Table 3 shows the correspondences between the FHIR resource *MedicationOrder* and the Content ODPs PlannedClinicalProcess and ClinicalProcess. In FHIR, a *MedicationOrder* is an order for both supply of the medication and the instructions for administration of the medicine to a patient. Within the ontology, it is reinterpreted as a plan (information entity) resulting from a prescription process (*MedicationPrescription*), which has as parts supply (*SupplyMedicationOrder*) and administration (*MedicationAdministrationOrder*) orders, information entities that have as realizable *MedicationDispense* and *MedicationAdministration* processes respectively.

In total, we have created 289 classes, 718 logical axioms and 119 object properties within the ontology. The DL expressivity is SIQ(D). Table 4 shows the OWL DL correspondences for the RDF predicates used.

BTL2 allows standardizing the way the ontology is queried, and content ODPs guide the building of the queries. Besides, BTL2 allows querying homogeneously different ontologies (e.g. SNOMED CT + FHIR), previously harmonized, as well as heterogeneous FHIR resources, now semantically related within the ontology. The

following query example (Figure 3) detects cases in which although the patient has an allergic intolerance to *sct:Ibuprofen*, it was prescribed to him. Additionally, logical reasoning supports more generic queries e.g. for products that contain ibuprofen.

Table 3. Correspondences between FHIR *MedicationOrder* and the Content ODPs PlannedClinicalProcess and ClinicalProcess. For the classes defined within the corresponding pattern, subclasses are introduced (e.g. *MedicationPrescription* **rdfs:subClassOf** *Process*). Predicates have been renamed for the use case and are equivalent to the ones defined within the corresponding pattern triple. Prefix ("fhir") has been omitted. The number indicates the pattern triple as described in Figure 2.

FHIR Resource	RDF Content ODP based representation
identifier	(2) *MedicationOrder* MEDICATIONORDERIDENTIFIER *PrescriptionOrderID*
dateWritten	(4) *MedicationPrescription* MEDICATIONORDERDATEWRITTEN *PrescriptionDateWritten*
status	(2) *MedicationOrder* MEDICATIONORDERSTATUS *PrescriptionOrderStatus*
dateEnded	(4) *MedicationPrescription* MEDICATIONORDERDATEENDED *PrescriptionDateEnded*
reasonEnded	(2) *MedicationOrder* MEDICATIONORDERREASONENDED *PrescriptionOrderReasonEnded*
patient	(6) *MedicationPrescription* MEDICATIONORDERPATIENT *Patient*
prescriber	(6) *MedicationPrescription* MEDICATIONORDERPRESCRIBER *Practicioner*
encounter	(1) *MedicationOrder* MEDICATIONORDERENCOUNTER *Encounter*
reason	(2) *MedicationOrder* MEDICATIONORDERREASON *PrescriptionOrderReason*
note	(2) *MedicationOrder* MEDICATIONORDERNOTE *PrescriptionOrderNote*
medication	(6) *MedicationAdministration* MEDICATIONORDERMEDICATION *PharmaceuticalProduct*
priorPrescription	(10) *MedicationPrescription* MEDICATIONORDERPRIORPRESCRIPTION *MedicationPrescription*
dosageInstruction	(2) *MedicationOrder* MEDICATIONORDERDOSAGEINSTRUCTION *MedicationAdministrationOrder*
dispenseRequest	(2) *MedicationOrder* MEDICATIONORDERDISPENSEREQUEST *SupplyMedicationOrder*
Substitution	*Not modelled*

Table 4. Examples of correspondence between RDF predicates and their OWL DL representations

RDF Predicate	OWL DL correspondence
MEDICATIONORDERPRESCRIBER	*fhir:MedicationPrescription* and **btl2:hasAgent** some (*fhir:Practicioner* and **btl2:isBearerOf** some *fhir:PrescriberRole*)
MEDICATIONORDERENCOUNTER	*fhir:MedicationOrder* and **btl2:isOutcomeOf** some *fhir:Encounter* and **btl2:isOutcomeOf** max 1 *fhir:Encounter*
MEDICATIONORDERMEDICATION	*fhir:MedicationAdminsitration* and **btl2:hasParticipant** some *fhir:PharmaceuticalProduct* and **btl2:hasParticipant** max 1 *fhir:PharmaceuticalProduct*

```
SELECT ?IbuprofenPrescription ?IbuprofenAllergy
WHERE {
  ?IbuprofenAllergy rdf:type fhir:AllergicIntolerance .
  ?IbuprofenAllergy btl2:hasRealization ?ProcessX .
  ?ProcessX btl2:hasParticipant ?AllergicPatientX .
  ?AllergicPatientX rdf:type Patient .
  ?ProcessX btl2:isCausedBy ?IbuprofenSubstance .
  ?IbuprofenSubstance rdf:type sct:Ibuprofen .
  ?IbuprofenPrescription rdf:type fhir:MedicationOrder .
  ?IbuprofenPrescription btl2:hasPart ?IbuprofenAdministrationOrder .
  ?IbuprofenAdministrationOrder rdf:type fhir:MedicationAdministrationOrder .
  ?IbuprofenAdministrationOrder btl2:hasRealization ?MedicationAdministrationX .
  ?MedicationAdministrationX rdf:type fhir:MedicationAdministration .
  ?MedicationAdministrationX btl2:hasParticipant ?ProductY .
  ?ProductY btl2:hasPart ?IbuprofenSubstance .
  ?IbuprofenPrescription btl2:isOutcomeOf ?EncounterX
  ?EncounterX rdf:type fhir:Encounter .
  ?EncounterX btl2:hasParticipant ?PatientX .}
```

Figure 3. SPARQL query example

4. Discussion and Conclusion

We have proposed to reinterpret FHIR resources by using BTL2 and Content ODPs. Three domain medication resources have been reinterpreted. We describe it for the *MedicationOrder* resource, in RDF and OWL DL and comment on the benefits.

At the moment the translation into the proposed RDF representation is a manual process. We are working on representing Content ODPs using shapes represented in SheX [17] and SHACL [18], in order to automate the translation and perform RDF graph data validation, supporting inference in cases based on both the open and closed world assumption [15].

FHIR resources allow an extension mechanism to add new attributes to the predefined list of resources, which can even modify the meaning of the resource [19] (e.g. not take a medication, as extension of MedicationOrder). A top-level ontology such as BTL2 aims standardizing this process and prevents semantic inconsistencies that risk semantic interoperability by creating silos of non-interoperable information.

Since RDF requires monotonicity (i.e. new assertions cannot invalidate old conclusions), the existing draft FHIR RDF representation focuses on EHR structure instead of content. Representing negation is therefore a critical point (e.g. patient does not have allergy to ibuprofen). For this and other representation issues such as elements ordering several approaches are possible, however out of the scope of this paper.

References

[1] Bender, D., Sartipi, K. HL7 FHIR: An Agile and RESTful approach to healthcare information exchange. In Proceedings IEEE International Symposium on Computer-Based Medical Systems. 2013 pp. 326-331
[2] Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The semantic web. Scientific american, 284(5), 28-37.
[3] HL7 FHIR Resource list: https://www.hl7.org/fhir/resourcelist.html Last accessed: November 2016
[4] The Yosemite Project. http://yosemiteproject.org Last accessed: November 2016-10-18
[5] HL7 and W3C WG. RDF for Semantic Interoperability. http://wiki.hl7.org/index.php?title=RDF_for_Semantic_Interoperability Last accessed:November 2016
[6] RDF FHIR Draft Representation. https://hl7-fhir.github.io/rdf Last accessed: November 2016
[7] Smith, B., Grenon, P. (2002). Basic formal ontology. http://ifomis.uni-saarland.de/bfo/
[8] Schulz S, Boeker M. BioTopLite: An Upper Level Ontology for the Life Sciences. Evolution, Design and Application. Informatik 2013. U. Furbach, S. Staab; editors(s). IOS Press; 2013
[9] Ontology for General Medical Science (OGMS). https://github.com/OGMS
[10] Schulz, S., and Ludger, J. "Formal ontologies in biomedical knowledge representation." *Yearb Med Inform* 8.1 (2013): 132-46.
[11] Schulz, S., Martínez-Costa, C. (2015). Harmonizing SNOMED CT with BioTopLite: An Exercise in Principled Ontology Alignment. Studies in health technology and informatics, 216, 832.
[12] SemanticHealthNet. http://www.semantichealthnet.eu/ Last accessed: November 2016
[13] Martínez-Costa, C., Schulz, S. Ontology content patterns as bridge for the semantic representation of clinical information Appl Clin Inf 2014; 5: 660–669
[14] Falbo, R. A., et al. Ontology patterns: clarifying concepts and terminology. CEUR-WS Vol 1188 14-26
[15] FHIR Ontology requirements. http://wiki.hl7.org/index.php?title=FHIR_Ontology_Requirements Last accessed:November 2016
[16] RDF Triples syntax: http://www.w3.org/TR/n-triples/ Last accessed:November 2016
[17] Prud'hommeaux, E., et al. (2014). Shape expressions: an RDF validation and transformation language. In Proceedings of the 10th International Conference on Semantic Systems (pp. 32-40). ACM.
[18] SHACL W3C Working Draft 2016 http://www.w3.org/TR/shacl/#ClosedConstraintComponent
[19] FHIR extensibility: https://www.hl7.org/fhir/extensibility.html Last accessed: November 2016

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-456

Communication of Children Symptoms in Emergency: Classification of the Terminology

Jessica ROCHAT[a,1] Johan SIEBERT[b] Annick GALETTO[b] Christian LOVIS[a] and Frédéric EHRLER[a]
[a] *Division of medical information sciences, University Hospitals of Geneva University of Geneva, Geneva, Switzerland*
[b] *Pediatric emergency department, University Hospitals of Geneva University of Geneva, Geneva, Switzerland*

Abstract. The significant part of non-urgent visits to the emergency highlight the necessity to advise people on the actions to take according to their symptoms. Although information sources are accessible through different channels their content often employs medical terminologies that are difficult to understand by laypersons. Our goal is to provide a terminology of the most common symptoms in pediatric emergency adapted to laypersons. This terminology is organized in a hierarchy by the mean of a card-sorting study. The resulting classification separates the symptoms into two main categories: "accident" and "illness" that are subdivided in 9 and 10 sub-categories. The study also revealed that some symptoms were not understood by the participants and had to be reformulated, confirming the importance of user-centered method. The classification resulting from this study will be evaluated through a tree-test.

Keywords. Health communication, access to information, consumer health information, terminology, symptoms, emergency health services, consumer behavior, consumer participation

1. Introduction

Many patients come to emergency for unnecessary reasons. Studies demonstrated that 30% of emergency department (ED) visits are in fact non-urgent [1]. The situation is similar in pediatric ED. This emphasizes the need to guide parents and help them to decide whether they should bring their children to emergency. Advices about the necessity to visit ED are usually given based on observed symptoms. The terminology employed to describe symptoms should be selected with care in order to be clearly understood by laypersons. Indeed, research in the field of consumer health vocabulary has demonstrated that consumers and health care professionals use different terminology to express themselves about health. This mismatch can hinder communication and health information seeking. In order to improve the communication

[1] Corresponding author : Division of medical information sciences, Geneva University Hospital, Rue Gabrielle-Perret-Gentil 4, 1205 Genève, Switzerland; Email: jessica.rochat@hcuge.ch

of health information to consumers, through mHealth, appropriate terminology must be constructed with their help [2, 3].

Others researches have attempted to build lists of the most common symptoms. We found three recent studies aiming to estimate the prevalence of most common symptoms. One study was based on a list of 25 symptoms [4]. Another study evaluates the prevalence of 23 different symptoms [5]. The most recent research studied the prevalence of 44 self-reported symptoms based on literature search [6]. Other studies identified the most common symptoms asking adults about their present symptoms or in the past 2, 4 or 6 weeks [7]. If all these results are very valuable, none of these researches concerned the symptoms that brought people to pediatric ED.

Due to the lack of appropriate terminology we aim at constructing a terminology of the most common symptoms of pediatric ED at the University Hospitals of Geneva (HUG).

2. Methods

The construction started with the collection of an initial list of symptoms. In a second stage, we decided to organize this list into a hierarchy since it facilitates the search and exploration [8]. To build the hierarchy we relied on a web based card sorting tool. This user-centered method allows eliciting categorization by end-users [9, 10]. This method allows us to identify the categories inside the hierarchy but also to verify that participants understand all the symptoms names.

2.1. Construction of the List of Symptoms

An initial list of symptoms, containing over 200 symptoms, has been collected from a variation of the Canadian triage scale [11] adapted for the Geneva pediatric ED. During a year, triage nurses of the ED selected the most commonly reported symptoms by patient in the pediatric ED and end up with a list of 47 symptoms of illness and accident.

2.2. The Card-Sorting Test

The card-sorting task requested participants to group items (the 47 symptoms) in coherent categories from their perspective and to label them. One category was initially provided to allow participants to regroup the items they did not understand.

2.3. Population

According to recommendations on the minimal sample size required to conduct card-sorting study we decided to recruit at least 30 participants [12]. Participants were recruited through social networks over a period of a month. The inclusion criteria were: be at least 18 years old and the exclusion criteria were: working in a medical field.

2.4. Results Analysis

Before analyzes, semantically similar categories but labeled differently were merged into standardized categories. For example, both categories called "skin problems" and "skin" were combined into a single standardized category called "skin problems".

To analyze the results, we used the Best Merge Method dendogram [13] provided by the software Optimal Workshop. This dendogram provides the proportion of participants that agree with each grouping.

We decided to keep only categories with at least 50% of participants' agreement. For symptoms for which no agreement was observed (less than 50%), we asked an ED physician to classify them within the categories established by the participants or to create new categories. We also discussed how to classify the symptoms that participants did not understand, and asked the physician to validate, or correct when necessary, the categorization made by the participants.

3. Results

The study took place from 18 August to 26 September 2016. The test has been completed by 35 participants. This sample included 30 women (86%) and 5 men (14%), 13 of them having children (37%) and 22 do not (63%).

On the 25 categories proposed by the dendogram, we kept 8 categories for which at least 50% of participants agreed. These categories are: 1) mouth / nose / ear / throat 2) digestion / intestine 3) urology / private parts 4) fever 5) skin problems 6) articulations / motricity 7) headache 8) accident.

Eight symptoms led to many disagreements: allergic reaction, depression/anxiety/crisis, cough/difficulty breathing, headache, bloody nose, oral thrush, hernia and whitlow. Four symptoms (oral thrush, whitlow, colic, hernia) were classified in the category "I do not know what that means" by up to 23% of participants.

As recommended, hierarchy should be limited to 8 items per level to provide effective navigation [8], therefore categories containing more than 8 items were split into subcategories. First, as participants created an "accident" category, we decided to also create an "illness" category to distinguish the two main types of symptoms in a first level. In the accident category, we created two sub-categories to group some similar symptoms and to avoid having too many items: 1) swallowed something/choked 2) sting. We also created the subcategory "rash" to reduce the number of items in the "skin problems" category. The same problem was avoided by creating the subcategory "mouth and throat".

Other changes were made following the discussion with the ED physician. Under the illness category, a "queasiness" category was created in order to insert the symptoms "queasiness without fever" and "queasiness with fever" previously located in the "fever" category. The category "mouth/nose/ear/throat" was lightly changed for "mouth/nose/ear/eyes" in order to reflect that the "eye" symptom also belong to the category. The symptoms, for which no agreement has been observed (less than 50%), were renamed when necessary and placed in the most suited existing category. Some symptoms were recognized as diagnostics and were renamed to match the associated symptoms (allergic reaction: "swollen lips/tongue" and "red patches and itching", hernia: "genital swelling" and "swelling in the groin", colic: "baby colic/crying crises", whitlow: "finger/nail infection", oral thrush: "oral thrush/white plates"). The three

symptoms that were not linked to any category were placed under the illness category (e.g. cough/difficulty breathing). Finally, composite symptoms composed of two sub-symptoms (e.g. rash with fever) were placed in several categories (e.g. “skin problems” and “fever”).

The final tree is presented in figure 1. Nodes in blue on the diagram are branches. Nodes in orange are leafs. In our case, the depth is of 4 levels and the breadth range from 2 to 10 nodes per level. The hierarchy is quite unbalanced since most of the nodes are regrouped under the illness category.

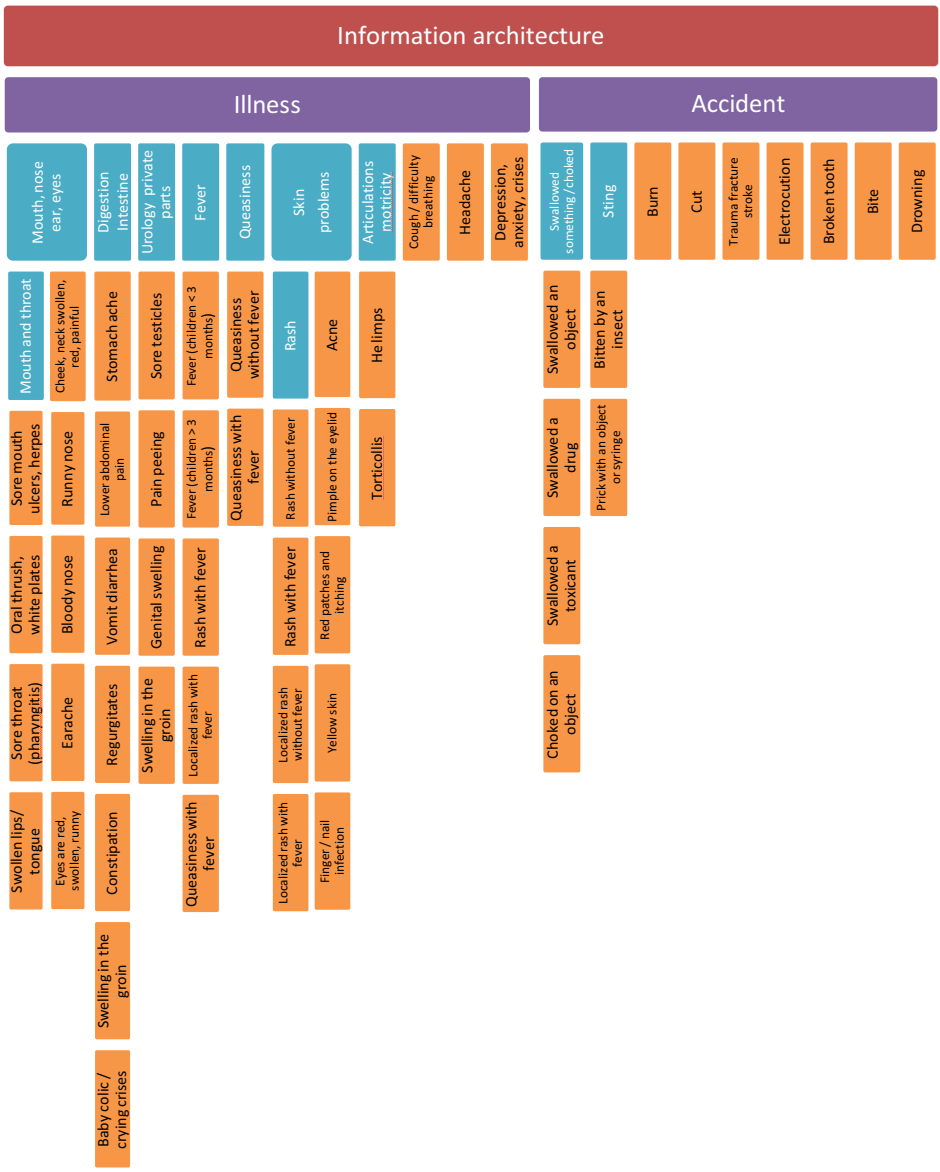


Figure 1. Terminology tree.

4. Discussion

Laypersons have difficulty to find information about their children's symptoms due to inappropriate terminology [14]. This underlines the importance of developing terminologies adapted to specific audience. A limitation of our study is that the initial list of symptoms is based on the Canadian triage scale. The ideal would have been to create a list of symptoms based on terms provided directly by laypersons who consult at the emergency department. Independently of the choice of the initial terminology, the organization of the symptoms made through card-sorting improves the findability of information provided to patients. Before implementing this terminology in a mobile application, the symptoms hierarchy will be tested through a tree-test in a further study to ensure its effectiveness.

References

[1] L. Uscher-Pines, J. Pines, A. Kellermann, E. Gillen, A. Mehrotra, Deciding to visit the emergency department for non-urgent conditions: a systematic review of the literature. *The American journal of managed care* **19** (2013), 47.
[2] Q.T. Zeng, T. Tse, Exploring and developing consumer health vocabularies. *Journal of the American Medical Informatics Association*, 13 (2006), 24-29.
[3] C.A. Smith, P.Z Stavri, Consumer health vocabulary. *Consumer Health Informatics*. Springer New York (2005), 122-128.
[4] A. McAteer, A.M. Elliott, P.C. Hannaford, Ascertaining the size of the symptom iceberg in a UK-wide community-based survey. *Br J Gen Pract* **61** (2011), e1-e11.
[5] D. Bruusgaard, H. Tschudi-Madsen, C. Ihlebaek, Y. Kamaleri, B. Natvig, Symptom load and functional status: results from the Ullensaker population study. *BMC public health* **12** (2012), 1085.
[6] S. Elnegaard, R.S. Andersen, A.F. Pedersen, P.V. Larsen, S. Rasmussen, D.E. Jarbol, Self-reported symptoms and healthcare seeking in the general population-exploring "The Symptom Iceberg". *BMC public health* **15** (2015), 685.
[7] L.M. Verbrugge, F.J. Ascione, Exploring the iceberg: common symptoms and how people care for them. *Medical care* **25** (1987), 539-569.
[8] A. Geven, R. Sefelin, M. Tscheligi, Depth and breadth away from the desktop: the optimal information hierarchy for mobile use. *Proceedings of the 8th conference on Human-computer interaction with mobile devices and services* (2006), 157-164.
[9] S. Bussolon, B. Russi, F.D. Missier, Online card sorting: as good as the paper version. *Proceedings of the 13th Eurpoean conference on Cognitive ergonomics: trust and control in complex socio-technical systems* (2006), 113-114.
[10] R. Dickstein, V. Mills, Usability testing at the University of Arizona Library: how to let the users in on the design. *Information technology and libraries* **19** (2000), 144.
[11] R. Beveridge, B. Clarke, L. Janes, N. Savage, J. Thompson, G. Dodd, A. Vadeboncoeur, L'échelle canadienne de triage & de gravité pour les départements d'urgence Guide d'implantation. *Can J Emerg Med* **1** (1999).
[12] T. Tullis, L. Wood, How many users are enough for a card-sorting study. *Proceedings UPA* (2004)
[13] Intepret the OptimalSort Dendrograms for open and hybrid card sorts. *Optimal Workshop* (2016). Available at: https://support.optimalworkshop.com/hc/en-us/articles/201997650-Intepret-the-OptimalSort-Dendrograms-for-open-and-hybrid-card-sorts.
[14] M. Benigeri, P. Pluye, Shortcomings of health information on the Internet. *Health promotion international* **18** (2003), 381-386.

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-461

SNOMED CT as Reference Terminology in the Danish National Home Care Documentation Standard

Kirstine Rosenbeck GØEG[a,1], Pia Britt ELBERG[a], Anne Randorff HØJEN[b] and Ulla Lund ESKILDSEN[c]
[a] *Department of Health Science and Technology, Aalborg University*
[b] *International Health Terminology Standardization Organization, IHTSDO*
[c] *Kommunernes Landsforening, Denmark*

Abstract. In Danish home care, multiple professions deliver services to citizens. FSIII is a national home care documentation standard, where one of the goals is to share documentation to improve coordination between these professional groups and avoid double documentation. The aim of this study was to develop a SNOMED-CT based navigation hierarchy to ensure that professions could preserve their documentation practice, to help avoid double documentation, and to ensure that the technical implementation did not require sophisticated semantic tools. The method involved mapping of non-SNOMED-CT content to SNOMED CT, visualization of merged graphs, identification of reference concepts, relating reference concepts to the documentation models of each profession, and representation of the navigation hierarchy in a reference set. The navigation hierarchy ensures that citizen conditions appear in a relevant context, regardless of which profession entered the data. Our approach paves the way for incremental standardization projects, where an implementation artefact, such as the navigation hierarchy, highlights the semantic features of SNOMED CT that can be used to reach specific business goals; in this case, sharing data across professional groups.

Keywords. SNOMED CT, navigation hierarchy, national documentation standard

1. Introduction

In Danish home care, multiple professions deliver services to citizens. For example, home nurses deliver health- and care-related services. Assistants assist with functional problems e.g. helps with tasks related to washing, cleaning and eating. Physiotherapists deliver rehabilitation and training. These different services are delivered as specified by two different Danish national acts (the health act and the service act) by the 98 Danish municipalities. IT-systems called "home care records" have been developed to digitally support these activities. However, the systems are characterized by a "single municipality, single stakeholder" scope and are largely unstructured, which makes it difficult to make municipality-care citizen-centered and track outcomes of new treatment initiatives, but also makes it difficult to use collected data for national secondary purposes such as statistics and research.

[1] Kirstine Rosenbeck Gøeg, Fr. Bajersvej 7C, DK-9220 Aalborg Ø, Denmark, E-mail: kirse@hst.aau.dk

In 2007, the idea of standardizing the documentation models of Danish home care was first formulated and it was adopted into the different Danish digitalization strategies in the years 2011-2013. In 2013 the project, FSIII (Cross Terminology in Municipalities, version three), was launched. Development of the common documentation model has been undertaken from 2013-2016, and implementation in Danish home care records is scheduled for 2017.

1.1. FSIII Documentation Model with Focus on Conditions

In current home care records, an unstructured to semi-structured description of the interventions that each professional group carry out, constitute the core part of the documentation following a visit. In FSIII, the point of departure will be citizen's health and functioning conditions. The idea is to focus care on the citizens' problems rather than the professional activities. In FSIII, it was decided to express a set of health conditions using SNOMED CT to support consistent recording of nursing related problems. This is described in our earlier work[1]. The health conditions are structured according to a national guideline for documenting nursing examinations with 12 nursing areas such as "nutrition", "skin and mucosa", "communication", "respiration and circulation" etc. These areas are well-recognized and used by the nursing profession in Denmark. However, from a semantic viewpoint the 12 areas are not well-defined; e.g. "respiration and circulations" are findings related to two entirely different body systems summarized in one nursing area.

Functioning conditions are used by assistants, and are recorded using a subset of ICF codes. ICF was chosen because the municipalities had already started using ICF to describe function. As part of the FSIII project, areas have been defined for the functional ability examination, so that assistants may structure their documentation under five areas e.g. "self-care abilities", "abilities related to carry out household chores", and "mental abilities".

In the years to come, additional condition sets are likely to be developed and included as part of FSIII, e.g. rehabilitation conditions as described by physiotherapists. In addition to conditions, a range of other attributes and value sets are also defined as part of the FSIII documentation model, e.g. patient demographics, interventions and intervention goals. However, any further description is omitted, because this paper focuses on the documentation of conditions, and how to handle the documentation overlap between professional groups.

1.2. Aim

The aim of this study is to suggest an approach to solve the following implementation challenges in the FSIII project:

- Different professional groups use different terminologies, and different models for documentation. Preserving documentation practices that support each professional group, while still standardizing information content, is a challenge.
- Different professional groups have overlapping documentation e.g. both assistants and nurses need to know about nutrition. As a result, each profession documents a citizen's nutrition related problems to plan their interventions. Coordination only happens on an ad hoc basis.

- The current clinical information system landscape is immature when it comes to SNOMED CT implementation [2]. Consequently, expecting any sophisticated terminology use, such as semantic querying, would not be realistic.

2. Methods

2.1. SNOMED CT Representation of Non-SNOMED CT Concepts

To ensure that assistants can continue using ICF, but still have a homogeneous terminological representation, we mapped each of the 30 ICF concepts to a SNOMED CT concept. We used a common set of mapping guidelines, and aimed for predictable retrieval properties rather than semantic precision[3]. The mapping was done by two annotators, and each disagreement was discussed and resolved. The mapping table was represented as a simple map reference set, as specified by IHTSDO [4].

2.2. Visualization of Sets and Identification of Reference Concepts

We made a common hierarchical graph for both sets of FSIII conditions, to provide an overview of the involved concepts. We visualized each concept and its super-type concepts from the health and functioning condition sets respectively, and merged the graphs as described in [5]. From the merged graph we could identify common ancestors, that preserved the clinical meaning of a group of health and functioning condition, and we named these concepts "reference concepts". Reference concepts can be understood as points of departure for predefined semantic queries i.e. from a clinical viewpoint it had to make sense to query for a reference concept and all its descendants within the FSIII condition sets. For example, it could make sense to query for all findings related to "activities of daily living" whether these findings were first documented by nurses or assistants. We ensured that each concept in the two condition sets, were subsumed by at least one reference concept, and if not, we added the concept itself as a reference concept.

2.3. Placing Reference Concepts in Each Professions Documentation Models

One objective of FSIII is to share information rather than document the same information twice. This means that nursing documentation should be informed by the documentation already made by assistants and vice versa. Consequently, it is not enough that nursing conditions are structured according to the 12 Danish nursing areas. Rather, *all* relevant FSIII conditions should be structured according to the 12 nursing areas and five functioning areas to ensure that regardless of e.g. nutrition being described by nurses or assistants, it is available when either group is looking for nutrition related information.
Ideally, it would be possible to obtain a meaningful reference concept for each of the 12 nursing areas and the five functioning areas. However, given that the areas are semantically poorly defined, we instead placed a group of reference concepts under each area i.e. each area was defined by a SNOMED CT expression of the following pattern:
$Area_x$=<<RC_1 OR <<RC_2……OR <<RC_n , *where RC is a reference concept.*

2.4. Representing the Navigation Hierarchy as an Ordered Type Reference Set

Given the overview in the merged graph, we could predict exactly which conditions each area-expression would retrieve. As such, distributing just the expressions would be adequate if mature SNOMED CT infrastructures were available. However, to accommodate immature systems, we developed a navigation hierarchy that expressed which reference concepts belonged to each area, and which health and functioning conditions belonged to each reference concept. We represented the navigation hierarchy as an ordered type reference set. Implementation of this reference set means that systems can utilize the pre-defined area expressions to avoid double documentation without implementing all of SNOMED CT or semantic querying.

3. Results

Of the 30 assistant conditions and 44 nursing conditions, there were 5 full matches between the two sets. However, the sets had substantial semantic overlap caused by one professional group documented in more detail compared to the other. This was especially evident for the findings related to functioning where nurses typically use coarse-granular concepts, such as 365178001|Finding related to ability to perform personal care activity|, whereas assistants would have each personal care activity specified e.g. 365180007 | Finding related to ability to perform washing and drying activities| and 365222007 | Finding related to ability to perform dressing activity|. Non-overlaps were mostly seen for health conditions where only nurses could take action e.g. different types of ulcers and pain. The merged graph (in Danish) is shown in [6].

The areas were defined using expressions. For example, the nursing area "Nutrition" is defined by:
$Area_{nutrition}$=*<<107647005 |Weight finding| OR <<284648005 |Dietary intake finding| OR <<116336009 | Eating / feeding / drinking finding|*
and the functioning area "Self-care ability" was defined by:
$Area_{self\ care}$= *<<365178001 |Finding related to ability to perform personal care activity| OR <<107647005 |Weight finding| OR <<284648005 |Dietary intake finding| OR <<116336009 | Eating / feeding / drinking finding| OR 130969003 | Health seeking behavior|*

The two presented expressions overlap meaning that if citizen information is related to any concepts in the three semantic groups related to nutrition i.e. <<107647005 |Weight finding|, <<284648005 |Dietary intake finding|, <<116336009 | Eating / feeding / drinking finding| this information should be shown when nurses try to get an overview of nutrition, and when assistants try to get an overview of self-care regardless of whom documented the information in the first place. Consequently, we can preserve the documentation practice of each professional group, while still viewing the information documented by other professional groups.

The navigation hierarchy was represented by a modified ordered type reference set, as specified by IHTSDO [4]. We replaced the referenced ComponentId and linkedId by sourceId, sourceRefset, destinationId and destinationRefset to allow the source and targets to come from either of the involved sets e.g. reference concepts, health conditions and functioning conditions because the whole idea was to relate the concepts of the different sets, which is outside the scope of a traditional ordered type reference set. The

navigation hierarchy, as well as the other reference sets and classifications related to FSIII are distributed to care record vendors by Local Government Denmark.

4. Discussion

In this study, we developed a navigation hierarchy that supports utilization of SNOMED CT as a reference terminology to bridge documentation of a citizen's conditions when entered by different professional groups using different terminologies and classifications. Whereas others have described the development of SNOMED CT reference sets for local, national or international use e.g. [7-9], terminology artefacts to ease implementation of SNOMED CT are much less common. The developed navigation hierarchy is an example of such an artefact. In the future, systems that use SNOMED CT artefacts might need to invest in sophisticated terminology management systems that allows mapping between classifications and semantic querying, and such systems are emerging; see e.g. the description of different solutions in [10]. However, most current clinical systems only allow very simple use of SNOMED CT. While waiting for more semantically sophisticated systems, national development towards semantic interoperability and better utilization of health related data continue. The developed navigation hierarchy demonstrates a pragmatic solution given international recommendations of incremental standardization e.g.[11]. Future studies await implementation projects by Danish vendors in 2017, to evaluate how the SNOMED CT based artefacts are adapted and used in the Danish home care records.

References

[1] Højen AR, Gøeg KR, Elberg PB. Re-use of SNOMED CT subset in development of the Danish national standard for home care nursing problems. Studies in health technology and informatics 2014;210:140-144.
[2] Lee D, Cornet R, Lau F, De Keizer N. A survey of SNOMED CT implementations. Journal of Biomedical Informatics 2012.
[3] Højen AR, Gøeg KR. SNOMED CT Implementation. Mapping Guidelines Facilitating Reuse of Data. Methods of information in medicine 2011;50(5):472-478.
[4] IHTSDO. SNOMED Clinical Terms. Technical Implementation Guide. 2010; Available at: http://ihtsdo.org/fileadmin/user_upload/doc/.
[5] Højen AR, Sundvall E, Gøeg KR. Methods and applications for visualization of SNOMED CT concept sets. Applied Clinical Informatics 2014;5(1):127-152.
[6] Kommunernes Landsforening. FSIII Referencebegreber og navigationshieraki v1.0. 2016; Available at: http://fs3.nu/filer/Dokumenter/Baggrundsmateriale/FSIII%20Referencebegreber%20og%20navigationshieraki.pdf.
[7] Francis Y. Lau, Raymond Simkus and Dennis Lee. A Methodology for Encoding Problem Lists with SNOMED CT in General Practice. KR-MED; 2008.
[8] Oluoch T, de Keizer N, Langat P, Alaska I, Ochieng K, Okeyo N, et al. A structured approach to recording AIDS-defining illnesses in Kenya: A SNOMED CT based solution. Journal of Biomedical Informatics 2015;56:387-394.
[9] Matney SA, Warren JJ, Evans JL, Kim TY, Coenen A, Auld VA. Development of the nursing problem list subset of SNOMED CT®. Journal of Biomedical Informatics 2012;45(4):683-688.
[10] Pathak J, Solbrig HR, Buntrock JD, Johnson TM, Chute CG. LexGrid: a framework for representing, storing, and querying biomedical terminologies from simple to sublime. Journal of the American Medical Informatics Association 2009;16(3):305-315.
[11] Kalra D, Vander Stichele R, Schulz S, Karlsson D, Gøeg KR, Cornet R, et al. D4.4 Policy and strategy recommendations – final report. Available at: http://assess-ct.eu/fileadmin/assess_ct/deliverables/assess_ct_d4.4.pdf.

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-466

Piloting a Collaborative Web-Based System for Testing ICD-11

Marc DONADA[a], Nenad KOSTANJSEK[b] ,Vincenzo DELLA MEA[a,1], Can CELIK[b], Robert JAKOB[b]
[a]*Dept. of Mathematics,Computer Science and Physics, University of Udine, Italy*
[b]*Classifications,Terminologies and Standards, World Health Organization, Geneva, Switzerland*

Abstract. Background: The 11th revision of the International Classification of Diseases (ICD-11), for the first time in ICD history, deployed web-based collaboration of experts and ICT tools. To ensure that ICD-11 is working well, it needs to be systematically field tested in different settings, across the world. This will be done by means of a number of experiments. In order to support its implementation, a web-based system (ICDfit) has been designed and developed. The present paper illustrates the current prototype of the system and its technical testing. Methods: the system has been designed according to WHO requirements, and implemented using PHP and MySQL. Then, a preliminary technical test has been designed and run in January 2016, involving 8 users. They had to carry out double coding, that is, coding case summaries with both ICD-10 and ICD-11, and answering quick questions on the coding difficulty. Results: the 8 users coded 632 cases each, spending an average of 163 seconds per case. While we found an issue in the mechanism used to record coding times, no further issues were found. Conclusion: the proposed system seems to be technically adequate for supporting future ICD-11 testing.

Keywords. International Classification of Diseases, Clinical Coding, Questionnaires

1. Introduction

The 11th revision of the International Classification of Diseases (ICD-11) has been recently presented to Member States for comment [1,2]. For the first time in the history of ICD revisions web-based expert collaboration and ICT tools were used in the development of the classification [3]. The ongoing ICD revision process is responding to numerous demands to align the classification to the latest scientific evidence and user requirements for more purpose-driven and IT compatible capturing and processing of diagnostic information.

To ensure that ICD-11 is working well, it needs to be systematically tested in different settings, across the world. This will be done by means of a number of testing arrangements.

The ICD-11 testing is unprecedented in the ICD revision history as in previous ICD revisions field testing was limited in scope (i.e. international field testing ICD-10

[1] Corresponding Author, Department of Mathematics, Computer Science and Physics, University of Udine, via delle Scienze 206, 33100 Udine, Italy. Email: vincenzo.dellamea@uniud.it

Chapter V) or conducted as an after step to facilitate the transition between the old and the new classification system (i.e. national Bridge Coding studies between ICD-9 and ICD-10) [4,5].

The overarching objective of the ICD-11 testing is to ensure systematic testing of ICD-11 before its use to increase consistency, accuracy and usability for morbidity and mortality coding as well as ascertain the comparability between ICD-10 and ICD-11.

To warrant strong and organized participation in the field testing from around the world WHO will designate ICD-11 Field Test Centers (FTC), which in turn will manage a network of national sites participating in the test. WHO Collaborating Centers, ICD-11 Topic Advisory Groups or other organizations with sufficient implementation capacity can serve as FTC.

In order to support the testing, a web-based system (ICDfit) has been designed and developed. To ensure functionality with large number of users an ICDfit prototype is currently undergoing a pilot testing. The paper describes features of the current ICDfit prototype and the process of its pilot testing.

2. System Requirements and Design

The system should implement the organization structure of field tests (FT) as foreseen by WHO, based on WHO coordination, FT centers (FTC) and FT sites (FTS). Two main kinds of studies are foreseen: one is devoted to line and case coding for mortality and morbidity using ICD-11 codes, or using both ICD-10 and ICD-11 codes. The diagnostic terms or statements used for line or case coding are extracted from clinical records, death certificates, etc. Another kind of study is devoted to asking basic questions on the updated classification to stakeholders, i.e., policy makers, classification experts, etc.

Users can belong to different categories: FTC Coordinator, FTS coordinator, Rater and Key Informant. Users have the following roles:

- The **Rater** carries out the basic work of participating into coding studies as classification user (i.e., coder). He/she accesses personalized ICD-FiT web page and checks for assigned cases in field test studies, completes case related forms for the respective study, fills in the evaluation form after completing all cases of a study.
- The **Key informant** participates into basic questions studies as classifications stakeholder. He/she accesses personalized ICD-FiT web page and fills in the forms related to the basic questions.
- An **FTS Coordinator** is able to invite and manage raters, and assign cases to them for each study in which the site is involved.
- An **FTC Coordinator** has the ability to assign new field test sites in their country, invite and assign key informants for basic questions studies, and assign cases to raters for each study in which the site is involved. This user role also has responsibility of translation of case summaries in local languages.
- The **WHO Coordinator** is able to insert new field test centers in the system, assign them to field test studies as well as inserting cases.

Collected data should be exported in CSV and Excel format for further processing.

Finally, the system should be sufficiently flexible to be adapted to new studies if needed.

Tools for web-based data collection exist, ranging from simple survey tools (like SurveyMonkey, Google Forms, etc) to complex systems addressing data collection in clinical trials (e.g., OpenClinica, REDCap, etc) [6]. However, due to the specific hierarchical organization of the field testing, together with the multilingual support needed also at the level of case summaries, we decided to implement an *ad-hoc* system specialized for the above mentioned requirements.

The system has been developed using PHP5, MySQL, Apache, on a Linux server. The three main FT instruments have been developed in both Western and Traditional Medicine version. An additional variant of the coding instrument, called "line coding", has been implemented to facilitate quick response.

Multilingual support has been developed, so that the system is currently available in 6 languages (although cases are only partially translated). The web interface has been developed using a responsive template in order to be accessible from tablets and smartphones too. Figure 1 shows an example of the web interface for double coding: on top there is a short case summary that has to be coded by raters, then there is room for up to 3 ICD-11 codes and the survey questions. After the end of ICD-11 coding, clicking on the "Start ICD-10 coding" will reveal an identical form section for ICD-10.

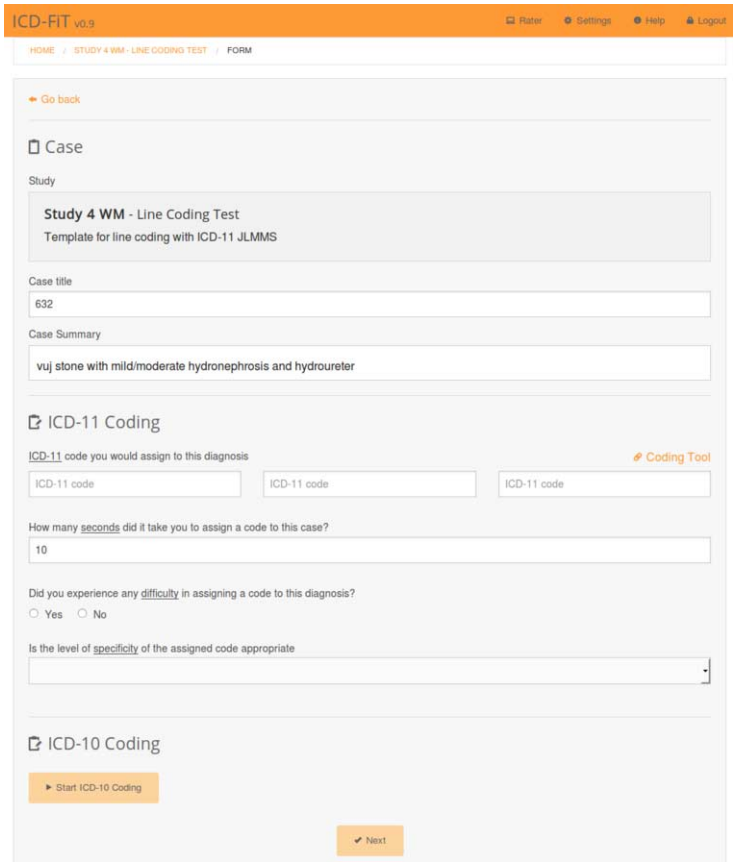


Figure 1. Line coding form.

3. Preliminary Technical Testing

In order to verify the functionality of the system, with also some preliminary indication on time needed for its usage and possible issues in the implementation, a technical test has been designed with the following features:

- 632 case summaries in English language have been prepared by the WHO coordinator and inserted into the system;
- A line coding instrument has been developed inside ICDfit to allow double coding using ICD-10 and ICD-11. Up to three codes per case summary were allowed for both ICD-10 and ICD-11 (fig.1);
- 8 coders from the Korean WHO-FIC collaboration center have been involved in coding the case summaries;
- The tool suggested for selecting codes in ICD-10 was the official ICD-10 browser, while for ICD-11 the new, experimental coding tool has been selected;
- Users were asked to code first using ICD-11, then using ICD-10.

Data collected included codes (up to 3 per case and per classification), answers to 3 questions per classification (regarding coding difficulties and code specificity), and time information (total time per case, time for ICD-10 coding, time for ICD-11 coding). A separate log of actions carried out on the system has also been maintained, in order to investigate on possible errors and issues.

The system has been integrated with available tools developed at WHO, in particular the ICD-10 browser and the ICD-11 coding tool.

4. Results

The technical testing experiment has been run in January 2016 and was successfully completed in two weeks by all coders, for a total of 5056 coded case summaries.

All cases were coded with at least one code, 153 had a second code in ICD-11 and 96 in ICD-10 by at least one coder, only 12 had a third code in ICD-11 vs. 6 in ICD-10.

The average time needed for double-coding a case and answering the survey questions was 163 seconds (range: 14-1426). However, when looking at the individual times for coding with just one classification, we found some discrepancies in the time spent. In fact, the average time for ICD-11 was 38 seconds, and for ICD-10 was 29 seconds.

An interview with users that revealed these discrepancies allowed to find that they were using a coding method that was more effective to them, but circumvented the time measuring method adopted in the system. In fact, some users maintained 3 browser tabs open with ICDFit, ICD-10 browser and ICD-11 coding tool: after having read the case summary, they individuated both codes on the respective tools, and then pasted them into the line coding tool.

Regarding the survey questions, coding difficulty was slightly more for ICD-11, as well as the number of answers indicating too much or too broad detail in ICD-11 codes. This is partly due to the not yet completed status of ICD-11. However, it should be noted that ICD-11 evaluation was not the aim of the test, more oriented towards technical functionalities of the system.

A thorough examination of server logs revealed relatively few problems. Some of them were related to network issues at very specific times; some other to session timeout due to inactivity or unstable network, that caused users to login again to complete their work. Session length was set at 24 minutes.

5. Discussion and Conclusions

The users participating in the technical testing were able to complete the assigned case summaries coding exercise in a relatively short time. The experiment allowed to discover some issues in the automatic time computation, due to the different behavior patterns users had in coding. This way, the apparent time needed for coding, as measured by the input timeline, was really short, although the total time resulting from logs was correct. Since it is almost impossible to technically constrain a more sequential behavior, if there is a real need for a precise evaluation of coding time, strict instructions should be provided to raters for the participation in the study.

However, the shortcuts identified by raters also suggest ways to make the overall field test process quicker, which is by itself a good result.

The integration with WHO tools has been also proven to function well, in particular for the newly released ICD-11 coding tool, that allows for natural language queries.

The overall time needed to double code a case is relatively short and sustainable, even in view of the tenths thousands answers needed for field test. However, before fully testing ICD-11, smaller tests will be made on selected, crucial chapters of the classification.

ICDfit has been adequate for helping to document and measure the code assignment in ICD-11 and ICD-10. Due to its generic features, it could also be used for other classifications. Further work will include mainly the development and implementation of a metrics to compare raters coding with gold standard coding, based on their distance on the classification tree, to be used for both ICD-10 and ICD-11.

References

[1] World Health Organization, “The International Classification of Diseases 11th Revision is due by 2018”, http://www.who.int/classifications/icd/revision/en/ (last accessed 10 Mar 2016).
[2] Boerma T, Harrison J, Jakob R, Mathers C, Schmider A, Weber S. Revising the ICD: explaining the WHO approach. Lancet. 2016 Oct 14.
[3] T.Tudorache, S.Falconer, C.Nyulas, M.A.Storey, T.B.Ustün, M.A.Musen, “Supporting the Collaborative Authoring of ICD-11 with WebProtégé”, Proc. of AMIA Annu Fall Symp. 2010: 802-6.
[4] Stieglitz RD, Zaudig M, Freyberger HJ, Dittmann V. Feasibility, suitability, and interrater reliability of ICD-10 during different stages of the ICD-10 field trial. Pharmacopsychiatry. 1990;23 Suppl 4:188-91
[5] Ustün TB, Goldberg D, Cooper J, Simon GE, Sartorius N. New classification for mental disorders with management guidelines for use in primary care: ICD-10 PHC chapter five. Br J Gen Pract. 1995 Apr;45(393):211-5.
[6] Franklin JD, Guidry A, Brinkley JF. A partnership approach for Electronic Data Capture in small-scale clinical trials. J Biomed Inform. 2011 Dec;44 Suppl 1:S103-8.

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-471

A Terminology in General Practice / Family Medicine to Represent Non-Clinical Aspects for Various Usages: The Q-Codes

Marc JAMOULLE[a,1] , Julien GROSJEAN[b], Melissa RESNICK[c], Ashwin ITTOO[d], Arthur TREUHERZ[e], Robert VANDER STICHELE[f], Elena CARDILLO[g], Stéfan J. DARMONI[b], Frank S. SHAMENEK[h] and Marc VANMEERBEEK[a]

[a] *Department of General Practice, University of Liège, Belgium*
[b] *Département d'Information et d'Informatique Médicale, University of Rouen, France*
[c] *University of Texas, Health Science Center at Houston, TX USA*
[d] *HEC Management School, University of Liège, Belgium*
[e] *Department of Health Sciences Terminology, BIREME/PAHO/WHO, Sao Paulo, Brazil*
[f] *Heymans Institute of Pharmacology, University of Ghent, Belgium*
[g] *Institute of Informatics and Telematics, Rende, Italy*
[h] *Consultant, New York, NY USA*

Abstract. The hereby proposed terminology called "Q-Codes" can be defined as an extension of the International Classification of Primary Care (ICPC-2). It deals with non-clinical concepts that are relevant in General Practice/Family Medicine (GP/FM). This terminology is a good way to put an emphasis on underestimated topics such as Teaching, Patient issues or Ethics. It aims at indexing GP/FM documents such as congress abstracts and theses to get a more comprehensive view about the GP/FM domain. The 182 identified Q-Codes have been very precisely defined by a college of experts (physicians and terminologists) from twelve countries. The result is available on the Health Terminology/Ontology Portal (http://www.hetop.org/Q) and formatted in OWL-2 for further semantic considerations and will be used to index the 2016 WONCA World congress communications.

Keywords. Q-Codes, general practice, family medicine, ICPC, categorization, controlled vocabulary, qualitative analysis, abstracts

1. Introduction

The field of medicine is blessed with a rich array of terminologies that support structured documentation of clinical information, and storage and retrieval of research publications. Overarching resources are the International Classification of Diseases (ICD), the Medical Subject Headings (MeSH) and the SNOMED CT. For most of the specialized medical domains have been built proper nomenclatures and classifications (e.g., the Systematized Nomenclature of Pathology, SNOP for pathologists) [1].

General Practice / Family Medicine (GP/FM) is a peculiar domain characterized by a very broad scope and a large array of research methods, and encompassing both

[1] Corresponding author: Dr Marc Jamoulle, MD GP, Charleroi, Belgium. E-mail: marc.jamoulle@doct.ulg.ac.be

clinical and non-clinical issues. Clinical issues are those aspects pertaining to signs and symptoms, reasons for encounter, and processes and diagnoses covered by the International Classification of Primary Care (ICPC) [2].

The non-clinical content of GP/FM is not fully represented either in indexes of textbooks in GP/FM, or in specific classifications or terminologies. The existing representation suffers from a top-down approach which is not always reflecting the complexity of the discipline.

To tackle this lack of non-clinical content in vocabularies, we hereby propose a new terminology for pragmatic use in real-life situations which can be used to index specific content of GP/FM (communications, literature or event managerial content of the contact with patients). Thus this resource is complementary to the ICPC-2.

In this paper, we aimed to a) describe the methodology applied for the creation of a taxonomy, b) on this basis, to present a domain-specialized terminological resource called the "Q-Codes" to facilitate the indexing of GP/FM non-clinical content and c) show how this resource could support the information retrieval of specific bibliographic information.

By doing this, we aim to contribute defining the limits of GP/FM.

This is also a preliminary work to address the future necessity of the application of semantic web technology on the GP/FM domain [3]. This approach refers to an ontology building process [4].

2. Methods

The first step of the methodology had the aim to design the taxonomy by identifying relevant concepts in a compiled corpus that includes GP/FM texts. We have studied the concepts identified in hundreds of communications of GPs during congresses from a bottom-up approach. The relevant concepts belong to the fields that are focusing on GP/FM activities (e.g. teaching, ethics, or environmental hazard issues).

The second step was the development of a terminological resource for each category of the resulting taxonomy. This has been formalized by defining concepts, hierarchy and mappings relationships. Several methods and tools were used to perform this step: i) Cimino's standard set of desiderata was applied to build the terminology content and structure [5], ii) we highlighted each concept by relevant bibliographic citations as well as by linking them to BabelNet[2], DBpedia[3] and to other reference terminologies, iii) we relied on the HeTOP multi-lingual and multi-terminology portal to fulfill each conceptual content and to manage the translations of terms and their definitions.

The final step was to evaluate and discuss the conceptual content of each created code of the taxonomy by involving several experts from all over the world (twenty four GPs and terminology experts from ten countries).

2.1. Selection of Abstracts and Corpus Compilation

A total of 1,702 abstracts were selected from six sources in English or French: (i) Wonca Europe 2007 (n=998), (ii) Portuguese 18th national conference of family

[2] http://babelnet.org
[3] http://wiki.dbpedia.org

medicine, 2013 (n=128), (iii) Congrès Confédération des Généralistes Enseignants (CNGE) Clermont-Ferrand 2013 (n=205), (iv) CNGE Lille 2014 (n=289), (v) SwissFamilyDocs Zurich 2014 (n=45), and (vi) Belgian GP/FM research congress Brussels 2014 (n=37). These six sources were selected due to the ease of accessibility of their abstracts.

2.2. Qualitative Methods for Development of the Taxonomy

Data collection stemmed from analyzing abstract proceedings from the mentioned GP/FM conferences. These data were analyzed in a grounded theory approach. It involves construction of a hypothesis or discovery of concepts through data analysis [6].

2.3. Tools

HeTOP is the Health Terminology/Ontology Portal[4] which provides access to 60 main health terminologies in several different languages. Due to these multilingual terminologies, HeTOP is used for many purposes. It is very useful not only for translators, terminologists and ontologists, but also for physicians coding patient records and using services on demand, e.g. info buttons. Finally, and most important for this research, HeTOP assists in indexing resources on the Internet.

We relied on HeTOP for this study to: a) find similar concepts in other vocabularies; b) create the Q-Codes terminology; c) manage each concept (labels, synonyms, definitions); d) perform manual mappings and e) give access to the final Q-Codes terminology through a web site.

2.4. The Q-Codes terminology

The ambition was to create a terminology to represent the non-clinical activities of GP/FM, by extending the 17 chapters of the clinical classification ICPC with an 18th chapter, called Q-Codes ("Q" being a letter in the alphabet not yet used in ICPC).

Each term identified in the first step has been converted into a concept (a "Q-Code"). Each concept has received a definition explaining its conceptual value. The extension of the concept, i.e. its use in several other online databases, has been documented through a careful search in a set of online dictionaries and terminologies. For each Q-Code, a minimum of properties was fulfilled to define the concept: a) the Preferred Label, b) one or more definitions, c) synonyms and linguistic variants, d) a sample of pertinent articles to the understanding of the main subject and e) BabelNet and DBpedia unique Ids and f) relevant MeSH terms related to the Q-Codes were mapped. Those mappings can lead to ease querying bibliographic databases (e.g. PubMed) for each specific Q-Code but they are also an important way to evaluate and assess the quality of Q-Codes definitions and conceptual content.

With regards to implementation, the Q-Codes (concepts, properties and relations) were formalized and implemented in Web Ontology Language (OWL). Based on the Resource Description Framework (RDF) standard, OWL is a knowledge-representation language which is considered the *de facto* language for ontology implementation. This

[4] http://www.hetop.org

task was a rough OWL-2 export from HeTOP to Web Protégé[5] without any description logic.

3. Results

The complete analysis of 1,702 French/English abstracts lead to the construction of a taxonomy composed by 182 terms. This taxonomy was enhanced to a terminological level according to Cimino's desiderata [5] and thanks to a complete support of semantic web technologies (HeTOP, Web Protégé, etc.) The resulting terminology called the "Q-Codes" consists of 182 concepts divided among 8 domains.

Each Q-Code (concept) was contentiously defined and tagged with English terms and definitions. Bibliographic citations and external concept URIs have been added to ensure semantic extension and validity. Table 1 gives an overview of the main Q-Codes domains with included covered topics.

Table 1. Q-Codes domains overview

Q-Code domain	Label	Examples of covered topics
QC	Patient's category	age, gender issues, abuse
QD	Family doctor's issue	communication, clinical prevention, medico legal issues
QE	Medical ethics	bioethics, professional ethics, info ethics
QH	Planetary health	environmental health, biological hazards, nuclear hazards
QP	Patient issue	patient safety, patient centeredness, quality of health care
QR	Research	research methods, research tools, epidemiology of primary care
QS	Structure of practice	primary care setting, primary care provider, practice relationship
QT	Knowledge management	teaching, training, knowledge dissemination

Results of the final step are related to the translation of the labels and definitions of the Q-Codes by different general practitioners in their native languages which are: French, Spanish, Brazilian-Portuguese, Dutch, Turkish, Korean and Vietnamese. Two terminologists reviewed the translations for three of the languages. One terminologist validated the English translation, and the second terminologist validated the Spanish and Portuguese translations. The Q-Codes multi-lingual terminology is available on HeTOP at http://www.hetop.org/Q (authentication required as wicc/wiccdemo).

4. Discussion

To the best of our knowledge, this is the first attempt to expand the ICPC coding system with an extension for managerial issues, thus covering non-clinical content, with the intent to improve performance in information storage and retrieval for research purposes in this broad, eclectic, and underserved domain of medicine.

[5] http://webprotege.stanford.edu/

4.1. Implications for Practice

We expect that the creation of this terminological resource for indexing abstracts and for facilitating Medline searches for general practitioners, researchers and students in medicine will reduce loss of knowledge in the domain of GP/FM. In addition, through better indexing of the grey literature (congress abstracts, master's and doctoral theses), we hope to enhance the accessibility of research results of general practitioners.

4.2. Implications for Research

End-users are often not very well-versed in knowledge-representation formalisms, and it remains to be proven that our proposed terminology will help them in dealing with more complex systems, such as MeSH, to support their information storage and retrieval activities. Nevertheless, the Q-Codes base is aimed at several uses: a) online repository of knowledge specific to GP/FM in several languages; b) online PubMed linked bibliographic system easy to use for training in GP/FM; c) resource for online e-learning; d) resource for the analysis of content of congresses in GP/FM in joint usage with ICPC-2 and indexation of gray literature in GP/FM; e) automatic or semi-automatic congresses indexing system; f) linking GP/FM to the web of data and the linked data initiative. Further work could be conducted to enhance the Q-Codes formalism with ontology building as the current OWL-2 version is a rough export.

5. Conclusion

"Q-Codes" are a terminology of non-clinical subjects in GP/FM. This work is the result of a two-year cooperative project between participants from twelve countries and eight languages. The 182 concepts have been elaborated in a bottom-up approach, by retrieving the topics most frequently addressed by GPs when they met in congresses. This work is freely available online at http://www.hetop.org/Q. The data is also available in the OWL-2 language for future use in the semantic web. The product, user guide and e-learning are available at http://3cgp.docpatient.net.

References

[1] R. Cornet & N. de Keizer, Forty years of SNOMED: a literature review, *BMC Medical Informatics and Decision Making* **8** (2008), Suppl 1, S2.
[2] I. Okkes, M. Jamoulle, H. Lamberts & N. Bentzen, ICPC-2-E: the electronic version of ICPC2. Differences from the printed version and the consequences, *Family Practice* **17(2)** (2000), 101–107.
[3] M. Jamoulle, R.H. Vander Stichele, E. Cardillo, J. Roumier, J. Grosjean & S.J. Darmoni, Semantic Web and the Future of Health Care Data in Family Practice, *Merit Research Journal of Medicine and Medical Sciences* **3(12)** (2015), 586–594.
[4] M. Fernández-López & A. Gómez-Pérez, Overview and analysis of methodologies for building ontologies, *The Knowledge Engineering Review* **17(02)** (2002), 129–156.
[5] J.J. Cimino, Desiderata for controlled medical vocabularies in the twenty-first century, *Methods of Information in Medicine* **37(4-5)** (1998), 394–403.
[6] C. Faggiolani, Perceived Identity: applying Grounded Theory in Libraries, *JLIS.it* (2011)

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-476

Computable Information Governance Contracts

James CUNNINGHAM [a,1], Gary LEEMING [b] and John AINSWORTH [a]
[a] *The Health eResearch Centre and The Farr Institute, Division of Informatics, Imaging and Data Sciences, School of Health Sciences, The University of Manchester*
[b] *Greater Manchester Academic Health Science Network*

Abstract. The risks of relinquishing control of electronic healthcare data for re-use in research are mitigated by the use of data sharing agreements and information governance procedures. These exist as legal, or quasi-legal, textual documents exchanged between data owners. Their existence outside of the digital realm leads to a situation where breaches of an agreement can only be detected and acted on post-hoc. We introduce the design of a system of computable contracts, specified formally, that can enforce the rules of data sharing agreements within the bounds of electronic health care systems.

Keywords. Information Storage and Retrieval, Database Management Systems

1. Introduction

Historically medical records have been paper-based, with the UK only recently moving towards an electronic representation [1], mirroring efforts and trends in Europe, North America and Australia amongst others [2]. Whilst the increased benefits this gives to health care provision and large scale medical research are clear [3] a sense of ownership has grown-up around the physical, paper-based, records an their corresponding electronic form. Value exists to both the organization and the practitioner in maintaining sole ownership of patient data. Economic value is created through the retention of a patient and even within the UK's relatively unified National Health Service (NHS) there exists competition between healthcare providers [4]. The challenges surrounding the reuse and trustworthiness of healthcare data are well recognised [5]. Whilst the use of electronic healthcare data is crucial for research, its continued use for such purposes is grounded in the maintenance of strict ethical frameworks of control [6]. Situations in which these ethical frameworks are perceived as ineffectual can lead to a breakdown in public trust in EHR systems, and in turn will threaten to bring down tighter regulatory controls on the use of data, threatening its effective use moving forwards [7]. Information Governance (IG), the policies procedures and controls that ensure the correct use of data, is therefore a crucial aspect in the use and practice of the digitisation of healthcare data. Without correct IG procedures, that both bring confidence in data use and ensure that this use falls within existing regulatory frameworks, the future benefits of digitised healthcare data will be

[1] Corresponding Author, Email: james.a.cunningham@manchester.ac.uk

severely limited. The documented procedures and legal agreements that specify the correct ethics-legal use of electronic healthcare data are specified and exchanged as documents distinct from the digitised data that they pertain to. There has been no 'big bang' move from the paper to the digital realm whereby all systems were switched at once to a modern representation; the transformation has been slow and piecemeal – an inevitable consequence of the crucial role that patient records play in day-to-day care and the need to maintain the smooth running of healthcare system. As such there are aspects of healthcare that have not yet started the move to the digital, information governance agreements and their use being one such aspect of the domain.

In this paper we describe a framework for specifying information governance contracts in computable form. These computable information governance contracts allow data owners to precisely specify the conditions under which certain actions (such as querying and retrieval) can be performed on medical data. These contracts can then produce electronically signed warrants which grant permissions to users and organisations the right to use data for purposes controlled by the underlying contracts. By attaching these warrants to formal requests for data both the data controller and the data user are provided with evidence that can ensure the correct and verifiable delivery of data for the purpose that its provision was intended, although production and delivery of the data itself lies outside the scope of this framework.

2. Method

The system we describe is based at its core on the notion of **contracts** for specifying the scope of allowed behaviour in terms of data sharing within an EHR system and **warrants** that specify data that can be returned from a system for a particular data access request. Contracts are mapped from agreements and exchanged between data providers. For a given access request for data a **request** object is specified which describes who is requesting the data and for what purpose that data will be used. Given a request and a contract a warrant is produced that describes the data accessible by that user for that purpose from a given data source. This warrant is then provided by the user to the data owner and used to determine what data is produced for that user.

The system as described below is given as a series of functional data types specified using the semantics of the F# programming language. Briefly, a type has an identifying name and a series of constructors (separated by the | symbol) that are used to define a value of that type. A constructor can, if needed, be given a series of values or the types specified by the 'of' keyword and separated by `*`. The types used to construct the warrants and contracts used by the system are given in the following.

A `Purpose` describes the use to which data will be put. Initially we have identified the core categories of `PatientCare` – accessing data in order to directly influence the course of care of a patient, `Research` – secondary research use, and `Benchmarking` – general data auditing. Additionally a `Purpose` can be defined as being `Any`, which indicates that once data is received it will potentially be used for any purpose within any standard constraints.

```
type Purpose
| Any
```

```
| PatientCare
| Research
| Benchmarking
```

An `Action` describes something taking place. In the simple case an `Action` can be For a `Purpose` as described above. An `Action` can also be specified as taking place `Until` or `After` a given date and time. Further a `Choice` represents a means of grouping a list of `Action`s.

```
type Action
| None
| For of Purpose
| Until of DateTime * Action
| After of DateTime * Action
| Choice of Action list
```

The type `Decision` signifies a binary choice of `Allow` or `Disallow` and is used to build the underlying form of the `Contract` datatype.

```
type Decision = Allow | Disallow
```

A `Code` is used to explicitly identify an item of data that is being queried against and is designed to hold a value translatable into an underlying clinical coding system. At present the type `Code` simply holds a string to be matched against. A more complex type definition of a clinical code is possible but was not deemed necessary for this prototype implementation.

```
type Code = CodeValue of string
```

On a basic level a `Contract` associates an `Action`, a list of `Code`s or both with a `Decision`. This captures the essence of the `Contract`, which is to allow or disallow given actions or requests for types of clinical code. `Contract`s which specify multiple `Action`s are grouped using the `Or` constructor and finally a default decision is specified using the `Other` constructor, although in practice and in the absence of a specific value this will be presumed to be `Disallow`.

```
type Contract =
  ForAction of Action * Decision
| ForCodes of Code list * Decision
| ForActionCodes of Action * Code list * Decision
| Or of Contract * Contract
| Otherwise of Decision
```

A `Request` mirrors in part the form of a contract, but pertains to only a single request with the corresponding `Or` constructor missing.

```
type Request =
  RequestFor of Action
| RequestForCodes of Code list
| RequestForActionCodes of Action * Code list
```

Finally a `Warrant`, produced as the result of a Request being matched against a Contract specifies that the holder of the `Warrant` be allowed to perform a given Action against the data held by the holder of the corresponding Contract. The Warrants constructor allows for a list of individual Warrants to be held in the data type.

```
type Warrant =
  WarrantedFor of Action
| Warrants of Warrant list
```

The programmatic API for accessing the system essentially consists of a single function that takes a Contract and a Request and produces a corresponding Warrant.

```
type VerifyRequest = Contract -> Request -> Warrant
```

The semantics of the function implementation are such that if the `Action` of the request matches a corresponding Action of a contract with a given `Allow` decision type then a `Warrant` for that `Action` is returned. If the corresponding decision type is `Disallow` then a `Warrant` is produced but with an `Action` of `None`. Similarly the underlying function implementation matches requests for `Codes` in a similar way, but the method for deciding whether two different `Code` strings match can vary based on the underlying semantics of the given coding system.

3. Results

We demonstrated the applicability of the system described above by translating existing information governance agreements into this computable format. These contracts were taken from real-world examples of information governance agreements placed on users of data from the Salford Integrated Record [8]. In the following we give a description in plain language of what the governance contracts specified in terms of the allowable forms of request and then show the corresponding implementation of a Contract datatype for two such contracts.

Contract 1: Data agreements were already in place for access to data for research purposes. Additionally access to data for auditing purposes would be granted for a six-month period. The contact granted benchmarking access to users for this period and then reverted back to the original agreement for ongoing research use. The corresponding formally typed definition is given below.

```
let contract1 = Or (ForAction (Until ("03-01-2016")
                                (For Benchmarking)))
                   (ForAction (For Research)) Allow
```

Contract 2. A data agreement was put in place allowing research access to data pertaining to Diabetes or for Asthma. The underlying coding system used was the Read code system. The corresponding formally typed definition is given below.

```
let contract2 = ForActionCodes (For Research) ["CD10.."; "H33.."] Allow
```

The system was used to translate a series of such contracts into the underlying formal specification, demonstrating the general feasibility of the system for such use.

4. Discussion

The exemplar results outlined above demonstrate the feasibility of the system in terms of its ability to translate existing information governance agreements into a computational format. This ability removes the existing 'analogue gap' between electronic health records existing in the digital domain and information governance contracts that lie outside that domain and yet are used to enforce the use and control of digital health records. As it stands the system is in proof-of-concept stage, and whilst we have demonstrated its feasibility through the translation of existing real world agreements, fully integrating it into an existing system would serve to fully prove its suitability in the real world. Given that information governance has at its core an underlying legal foundation, there will likely be barriers to adoption in terms of a requirement for computational contracts as described here to be formally verified to meet such legal requirements. This, combined with the natural reticence that exists towards the adoption of novel methods to be applied to a field as sensitive as personal medical data may make moving forwards in the real world difficult.

Information governance is a vital aspect in the use of electronic healthcare data. Without strict information governance procedures in place the use of electronic health-care data, particularly for research purposes, will remain limited with respect to its potential use. By moving the specification of IG contracts into a computable form we have taken a step towards demonstrating the potential for formally enhancing the trust that can be placed in such systems. This is the first step in integrating the currently informal (in a computational sense) specification of IG procedures into the digital realm within which patient records already reside. Moving forwards we see this as a key future direction into the full digitisation of the healthcare domain, which we see as a crucial future direction of the field.

References

[1] House of Commons Health Committee, "The electronic patient record." http://www.parliament. the-stationery-office.co.uk/pa/cm200607/cmselect/cmhealth/422/422.pdf, July 2007. [Online; accessed 7-Nov-2016].

[2] A. Cornwall, "Electronic health records: An international perspective," *Health Issues*, no. 73, 2002.

[3] T. B. Murdoch and A. S. Detsky, "The inevitable application of big data to health care," *JAMA*, vol. 309, no. 13, pp. 1351–1352, 2013.

[4] R. Lewis, J. Smith, and A. Harrison, "From quasi-market to market in the National Health Service in England: what does this mean for the purchasing of health services?," *Journal of health services research & policy*, vol. 14, pp. 44–51, 2009.

[5] A. Geissbuhler, C. Safran, I. Buchan, R. Bellazzi, S. Labkoff, K. Eilenberg, A. Leese, C. Richardson, J. Mantas, P. Murray, *et al.*, "Trustworthy reuse of health data: a transnational perspective," *International journal of medical informatics*, vol. 82, no. 1, pp. 1–9, 2013.

[6] B. M. Knoppers and R. Chadwick, "The ethics weathervane," *BMC medical ethics*, vol. 16, no. 1, p. 58, 2015.

[7] M. J. Taylor and N. Taylor, "Health research access to personal confidential data in England and Wales: assessing any gap in public attitude between preferable and acceptable models of consent," *Life sciences, society and policy*, vol. 10, no. 1, p. 1, 2014.

[8] NHS Salford, "Salford integrated record. sharing patient information locally." http://www.salfordccg.nhs.uk/documents/Publications/SIRA5Booklet.pdf, July 2016. [Online; accessed 7-Nov-2016].

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-481

A Semantic Framework for Logical Cross-Validation, Evaluation and Impact Analyses of Population Health Interventions

Arash SHABAN-NEJAD[a1], Anya OKHMATOVSKAIA[b], Eun Kyong SHIN[a], Robert L. DAVIS[a], Brandi E. FRANKLIN[c], David L. BUCKERIDGE[b]
[a] *University of Tennessee Health Science Center – Oak-Ridge National Lab (UTHSC-ORNL) Center for Biomedical Informatics, Department of Pediatrics, Memphis, Tennessee, USA*
[b] *McGill Clinical & Health Informatics, Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, Quebec, CANADA*
[c] *Health Outcomes and Policy Research, Department of Clinical Pharmacy, The University of Tennessee Health Science Center, Memphis, Tennessee, USA*

Abstract. Most chronic diseases are a result of a complex web of causative and correlated factors. As a result, effective public health or clinical interventions that intend to generate a sustainable change in these diseases most often use a combination of strategies or programs. To optimize comparative effectiveness evaluations and select the most efficient intervention(s), stakeholders (i.e. public health institutions, policy-makers and advocacy groups, practitioners, insurers, clinicians, and researchers) need access to reliable assessment methods. Building on the theory of Evidence-Based Public Health (EBPH) we introduce a knowledge-based framework for evaluating the consistency and effectiveness of public health programs, interventions, and policies. We use a semantic inference model that assists decision-makers in finding inconsistencies, identifying selection and information biases, and with identifying confounding and hidden dependencies in different public health programs and interventions. The use of formal ontologies for automatic evaluation and assessment of public health programs improves program transparency to stakeholders and decision makers, which in turn increases buy-in and acceptance of methods, connects multiple evaluation activities, and strengthens cost analysis.

Keywords. Program Evaluation, Outcome assessment, Ontology, Inference, Public Health Intervention

1. Introduction

Chronic diseases impose a massive global burden [1]. To improve the planning and evaluation of disease prevention interventions and to assist multiple stakeholders and decision-makers in making correct and unbiased inferences about population health, several data and analytical issues should be addressed [2]. Major barriers to the planning

[1] Corresponding Author: Arash Shaban-Nejad, University of Tennessee Health Science Center – Oak-Ridge National Lab (UTHSC-ORNL) Center for Biomedical Informatics, Department of Pediatrics, 497R Le Bonheur Research Center, 50 N Dunlap St, Memphis TN 38103, USA, Tel: (901) 287-5366; Email: ashabann@uthsc.edu

or evaluation of effective public health interventions include the lack of reliable mechanisms to check the validity and precision of indicators and inability to identify errors, redundancies and inconsistencies (e.g. contradictory definitions such as having diabetic patients with no glucose metabolism disorder) in the coding of indicators [2, 3] that are used to evaluate the core components of an intervention. Large inconsistencies in indicator values (e.g. when different definitions and values are assigned to the same indicator) can generate confusion among policy makers and the public at large, and may lead to conflicting policy conclusions. Also many public health policies are not accurately defined, may not be cost-effective and important health outcomes are often challenging to measure and quantify against notions of the public good and safety. Defining components of interventions and policies with respect to their context and their level of development is a key prerequisite for designing interventions appropriate to meeting assessment requirements based on stakeholders' needs.

Evidence Based Public Health (EBPH) has been defined as the process of "integrating science-based interventions with community preferences to improve the health of populations" [4]. In the EBPH context, public health interventions and policies are evaluated based on their effectiveness as described in the best available evidence regarding existing guidelines and interventions for confronting the causes of a specific disease or condition in a population. The major goal of our study is to provide an automatic knowledge-based evaluation framework that aligns the results generated through semantic reasoning with the best available and reliable evidence. In turn, this will allow us to assess the relative effectiveness of various possible interventions, and support the logical evaluation of public health policies. To this end we have designed the POLicy EVAlUation & Logical Testing (*POLE.VAULT*) Framework to assist stakeholders and decision-makers in making informed decisions about different health-related interventions, programs and ultimately policies based on the contextual knowledge at various levels of granularity.

2. Methods & Results

Most existing program and policy assessment methods are either based on survey data or based on a set of pre-defined assumptions without considering the context and its dynamics. Using contextual knowledge captured in ontologies, we can analyze how context influences outcomes. Root cause analysis (RCA) [5] is a method for identifying the underlying causes for a public health problem that needs to be addressed and for proposing remedial actions. Using semantic inference along with the RCA principles, one can determine what elements in a disease causal pathway (e.g. behaviors, actions, or conditions) have to be changed to stop recurrence of similar, undesired outcomes. The main question for evaluating an intervention is whether it worked (e.g. changed a situation or behavior) in a target community. According to Bloom's taxonomy of learning[2], from knowledge to evaluation [6], to answer this question, the stakeholders (i.e. public health institutions, policy-makers and advocacy groups, practitioners, insurers, clinicians, and researchers) need to access knowledge to address lower-level questions such as:

- What motivated the need for change?

[2] Knowledge ⇒ Comprehension ⇒ Application ⇒ Analysis ⇒ Synthesis ⇒ Evaluation

- Are there resources available and adequate to implement *Intervention* I_1?
- What is the cultural, political and economic context in which change will take place?
- What types of activities lead to improved outcomes (e.g. improved health)?
- Which features of the intervention should be measured when evaluating its value/effectiveness? What evidence is needed to represent these features?

To address these types of questions we design a knowledge-based framework for intervention/policy evaluation at the community setting [7]. PopHR [8, 9] is a semantic web platform that uses ontologies to integrate multiple clinical and administrative data sources to provide a coherent view of the health of populations. Building on top of the PopHR semantic platform, we integrate the knowledge about causal and correlational evidence along with information about the target populations. By doing this integration of knowledge about interventions we can then formulate queries for identifying potential interventions, programs and evaluating their prospective outcomes. For example:

- What would be the impact of intervention I_1 on prevalence of *Disease* D_1 in three years from now?
- What are the relative impacts of *Interventions* $I_1, I_2, \ldots, I_n$ on increasing youth physical activity in school and community settings?
- Can *Intervention* I_x, which was successful in *Population* P_1, be used effectively for *Population* P_2, to prevent *Condition* C_y?

As shown in the activity diagram that demonstrates the *POLE.VAULT* framework (Figure 1) the automatic evaluation of interventions starts by assessing the interventions' logic models.

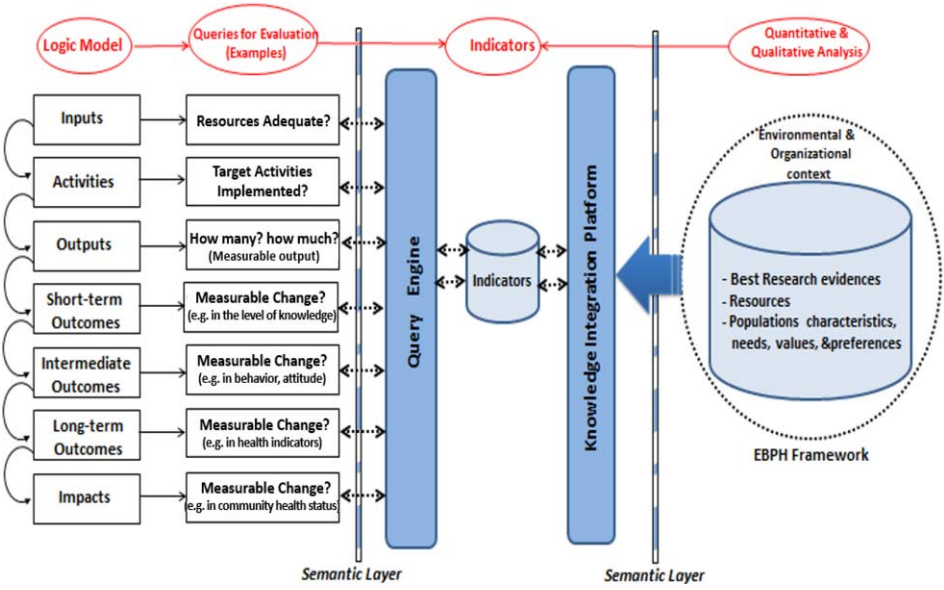


Figure 1. An abstract representation of the *POLE.VAULT* knowledge-based platform for logical cross-validation, and evaluation of population health interventions and programs.

Logic models are representations of the relationship between the resources, activities, outputs, outcomes and impacts of a public health program and therefore are important tools [10] in conducting program evaluations. This step is critical because, regardless of whether or not evaluations are process or outcome oriented, or qualitative or quantitative, they need to assess whether critical program components or activities were implemented and whether they had an impact on mediating outcomes, important behaviors and overall health goals. After populating our formal ontology model using the existing structured (e.g. databases) or unstructured (e.g. free text reports, surveys and interviews) data sources we compute different indicators to answer queries assessing each component of the logic model. For example, to check whether an intervention had the desired impact we define queries to examine changes in the population health status (positively, negatively, or unchanged). After this process is complete, the next step is aligning the logic model with the semantic knowledge platform and the conceptual model created using causal knowledge along with evidence on interventions' effectiveness (within semantic layers shown in Figure 1).

When integrating qualitative and quantitative data and information from different sources (some might even be contradictory) systematic errors would be inevitable. To ensure the consistency of our knowledge while evaluating interventions we perform logical cross-validation through a semantic reasoner. For this purpose and for logical assessment of the model and semantic query answering we use a Tableau-based description logic (DLs) [11] Reasoner. The alignment between the logic model and our evidence-based knowledge platform is attained through a set of service ontologies. As an example one might be interested to know whether sufficient resources exist to implement an online program to decrease Type II diabetes in Memphis TN by helping people to make positive changes in people's eating and exercise behavior? The high level conceptual model, demonstrated in Figure 2, shows different components for typical programs resource allocation, which are populated using causal evidence. After analyzing evidence on similar interventions and policies in other comparable settings (e.g. comparable populations sharing common features) the system provides an answer to the query above.

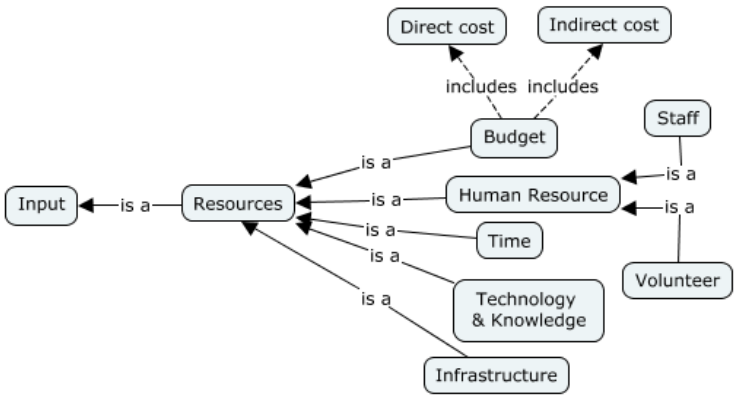


Figure 2. Partial representation of the high level conceptual model to represent "*Input"* in a Logic Model.

Using the formally represented knowledge and the logical reasoners/query engines *POLE.VAULT* also assists the stakeholders to evaluate (or forecast) various outcomes of

one specific intervention (or series of interventions) through logical inference and deductive querying (e.g. what are the possible outcomes of a specific intervention on the prevalence of diabetes in a defined community?)

As another example, if we removed blighted properties in neighborhoods 'x, y and z', what would be the impacts on health outcomes, such as asthma and hypertension as well as impacts on rates of crime, burglary and falls? Also, how are these health and non-health determinants and indicators related? Answering these questions allows different stakeholders to see what interventions are likely to give the biggest return on investment, and to enable them to make evidence based decisions accordingly.

3. Discussion

By assessing program outcomes through a set of indicators, the *POLE.VAULT* platform enables comparative analyses of the effectiveness and cost of different public health interventions. Moreover, in the future our platform intends to evaluate programs based on their coverage and effectiveness by highlighting the proportion of the defined objectives and goals that a program is capable of addressing, within a specific resource range (e.g. time period, budget bracket and human resources). Our future work also will be focused on extending the analytics power provided through the alignment between *POLE.VAULT* semantic framework and a statistical inference model to help shed light on potential causes of failure for unstable, fragile and uncertain programs and interventions. Finally, we will incorporate knowledge about how to effectively implement interventions, at which point the system will be able to recommend remedies to fix the problems in the failed interventions or suggest new interventions.

References

[1] World Health Organization. Global Status Report on Noncommunicable Diseases 2010. Geneva: 2011.
[2] R.G. Parrish, Measuring population health outcomes. *Prev Chronic Dis.* **7**(2010), A71.
[3] J. Datta, and M. Petticrew, Challenges to evaluating complex interventions: a content analysis of published papers. *BMC Public Health* **13** (2013), 568.
[4] N.D. Kohatsu, J.G. Robinson, J.C. Torner, Evidence-based public health: an evolving concept. *Am J Prev Med.* **27**(2004), 517–8421.
[5] T.O.A. Lehtinen, M.V. Mäntylä , J. Vanhanen, Development and evaluation of a lightweight root cause analysis method (ARCA method) – Field studies at four software companies. Information and Software Technology 2011, 53(10): 1045–1061.
[6] D.R. Clark, *Bloom's Taxonomy of Learning Domains*, 1999. Retrieved 15 May 2016.
[7] A. Bauman, and T.D. Koepsell, Epidemic issues in Community Interventions. In R.C. Brownson and D.B. Petitti (eds.), *Applied Epidemiology: Theory to Practice*. Oxford University Press, 2nd edition, 2006.
[8] A. Shaban-Nejad, M. Lavigne, A Okhmatovskaia, D.L. Buckeridge DL, PopHR: a knowledge-based platform to support integration, analysis, and visualization of population health data. *Ann N Y Acad Sci*. 2016 Oct 17. doi: 10.1111/nyas.13271. [Epub ahead of print].
[9] D.L. Buckeridge, M.T. Izadi, A. Shaban-Nejad, L. Mondor, C. Jauvin, L. Dubé, Y. Jang, R. Tamblyn, An infrastructure for real-time population health assessment and monitoring. *IBM Journal of Research and Development* **56** (2012): 2:1-11.
[10] J.A. McLaughlin, G.B. Jordan, Logic models: a tool for telling your programs performance story. *Evaluation and Program Planning* **22** (1999): 65–872.
[11] F. Baader, D. Calvanese, D.L. McGuinness, D. Nardi, and P.F. Patel-Schneider (editors), *The Description Logic Handbook*. Cambridge University Press, 2nd edition, 2007.

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-486

Discovering Central Practitioners in a Medical Discussion Forum Using Semantic Web Analytics

Enayat RAJABI[a,1] and Syed Sibte Raza ABIDI[a]
[a] *NICHE Research Group, Faculty of Computer Science, Dalhousie University, Halifax, NS, Canada*

Abstract. The aim of this paper is to investigate semantic web based methods to enrich and transform a medical discussion forum in order to perform semantics-driven social network analysis. We use the centrality measures as well as semantic similarity metrics to identify the most influential practitioners within a discussion forum. The centrality results of our approach are in line with centrality measures produced by traditional SNA methods, thus validating the applicability of semantic web based methods for SNA, particularly for analyzing social networks for specialized discussion forums.

Keywords. Medical forum, Semantic Web, Social Network Analysis, Centrality Measures.

1. Introduction

Social Network Analysis (SNA) is performed by analytical approaches that quantify an actor's influence, participation, reliability, respect and function within a given context or community [1]. In a medical online discussion forum, for example, the centrality measure provides insights into the influence of practitioners and physicians, whereas the cohesion measure informs how well the actors connect with each other as community [2]. Given the richness of social data, which extends beyond physical relationships between actors, there is an opportunity to investigate the semantics of these relationships to further understand the context and content of social networks. This brings to relief the application of semantic web technologies to semantically represent and analyze social networks. Semantic web technologies [3] can be utilized in the social network sphere to (a) enrich social data by aligning with external vocabularies, (b) categorize the interactions between actors in terms of a domain-specific model; (c) reason over the social data to derive knowledge-driven inferences and generalized models about the relationship roles and functions; and (d) link social data with external ontologies to infer new social relationships both within the social network, external social networks or semantically defined web resources. Previous studies have leveraged semantic web technologies to aggregate, extract, visualize and analyze different kinds of social data by making use of semantic ontologies and

[1] Corresponding author: 6050 University Avenue, PO BOX 15000, Halifax, NS B3H 4R2, Canada; E-mail: rajabi@dal.ca.

analytics [1] [4]. In this paper, we investigate semantic web-based methods to transform social network data to perform semantics-driven social network analysis to identify the central actors/practitioners in an online medical discussion forum. Our contention is that the central practitioners in an online discussion forum are relatively more active than the rest of community, and hence such central practitioners can serve as knowledge brokers in a knowledge translation context. This article is structured as follows. The proposed framework for semantic enrichment of social data and the social data used within our medical discussion forum is presented in Section 2. We will discuss the results of approach in Section 3, before concluding in Section 4.

2. Proposed Approach

In our work, we analyze social data collected from Surginet[2] which is an online medical discussion forum. The users of Surginet are general surgeons and medical practitioners in related disciplines who discuss various academic and clinical subjects specific to the specialty of General Surgery. The working of the medical discussion forum is as follows: (a) To initiate a topic-specific discussion, a practitioner poses a topic-specific question, comment or information item to the community; (b) In response, practitioners interested in the topic provide their response; (c) A topic-specific discussion thread starts to evolve which comprises a set of practitioners connected with each other, based on their response pattern. The discussion forum interactions are rendered as a 2-mode social network, where one mode is a set of practitioners and the other mode is a set of discussion threads—practitioners are connected with each other via the discussion thread they jointly participated in. Surginet comprised 231 practitioners and 2,111 discussion threads (comprising around 18,000 individual messages/posting by the practitioners) covering a range of topics related to surgery.

Our approach to semantically analyze medical social network data is as follows: (i) enrich social data generated within a medical discussion forum with semantic descriptions of the discussion topic using external domain-specific ontologies, (ii) expose the enriched social network data as a semantic web graph; (iii) analyze the graph to identify the active/influential practitioners (based on participation levels). Figure 1 illustrates our framework which has two main components: Social Data Transformation to semantic representation of Resource Description Framework (RDF) and Semantic Social Data Analysis, described as follows:

2.1. Social Data Transformation

This component takes as input traditional social data, represented in a relational database, and generates a corresponding RDF dataset based on a social data ontology. To identify the semantic similarity between the practitioners, we assigned MeSH terminologies[3] to each practitioner, derived from their posted messages using MetaMap tool[4]. The MeSH terms along with their frequency as used by each practitioner in each thread provide us a measurement to calculate the semantic closeness between the

[2] http://www3.sympatico.ca/tgilas/SURGINET.html

[3] https://www.nlm.nih.gov/mesh

[4] https://metamap.nlm.nih.gov

practitioners based on their message content. As the result of this step, a two-mode RDF dataset including practitioners' information along with associated MeSH terms was generated based on the social data ontology and a set of assigned MeSH terms.

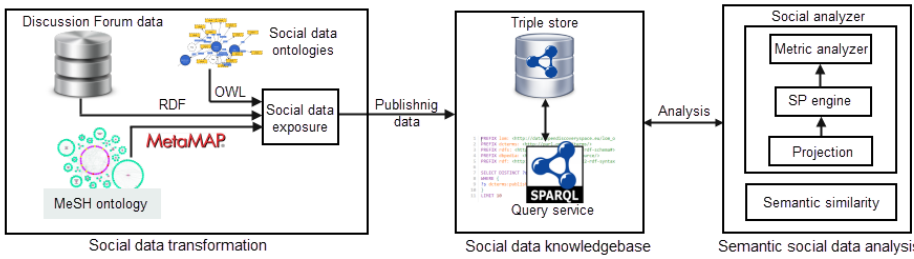


Figure 1. The process of exposing and analyzing of social data.

2.2. Semantic Social Data Analysis

To analyze the dataset and identify the central practitioners, we used a projection method [5] to map our two-mode dataset to a one-mode graph as described follows. We considered a large two-mode network modeled as a bipartite graph G = (, ⊥, E). The ⊥-projection of G is the graph G⊥ = (⊥, E⊥) in which two nodes (of ⊥) are linked together if they have at least one neighbor in common in G: E⊥ = {(u, v), ∃x ∈ : (u, x) ∈ E and (v, x) ∈ E}. A weighted one-mode network is created according to the two-mode network by defining the weights as the number of co-occurrences (e.g., the number of posts that two practitioners send to the same thread). To formally describe this method, we formalize it as: $W_{ij} = \sum_p 1$ where W_{ij} is the weight between node i and node j, and p is the nodes that connect node i and node j to each other (e.g., the threads in which two practitioners participate). Newman et. al [6] proposed to discount for the size of users interactions in a network by defining the weights among the nodes using the following formula: $W_{ij} = \sum_p \frac{1}{N_p - 1}$. To map the above formula on a discussion forum, N_p is the number of practitioners participated in thread p. In a discussion forum network, the above mentioned approach implies that if two practitioners who only post on a same thread with no other practitioners get a weight of 1. However, if the two practitioners post in a thread, which has another practitioner posted, the weight on the edge between them is 0.5 ($\frac{1}{(3-1)}$). As the result, the graph was converted to a one-mode RDF model using the above-mentioned approach. To measure the centrality of practitioners in a discussion forum, using traditional SNA approach we calculated degree centrality and closeness centrality metrics. We define the degree centrality in a discussion forum simply as the number of threads that practitioners participate and it is considered as the weights of a node in the graph. Closeness centrality also refers to the extent to which an individual can reach all others in the network in the fewest number of direct and indirect links. To calculate the closeness centrality metric, we utilize the Dijkstra algorithm to find shortest paths in graphs, which is performed by the Shortest Path engine (SP engine) of the framework. Typically, the closeness centrality of node "i" is calculated as: $closeness(i) = \sum_j \frac{1}{d_{ij}}$ where d_{ij} is the shortest path between node "i" and node "j". In a discussion forum, the closeness centrality between two practitioners implies how close they are to each other according to their posted messages in common threads. In other words, central

practitioners with high closeness centrality discussed relevant subjects with most of the practitioners in the discussion forum and thus have been more conceptually close to them than others. A semantic similarity metric usually takes as input two concepts, and returns a numeric score that quantifies how much they are alike, based on is-a relations. In our social graph, two practitioners are semantically considered as similar if they have been assigned by similar MeSH terms in the graph. To this aim, we followed the Kang et al. [7] approach where we calculate the MeSH terms similarity in a graph and discover similar practitioners who used the corresponding MeSH terms in the discussion forum. In this method, we used Jiang similarity metric [8] to discover the semantic similarity between central practitioners. We calculated the similarity following two steps:

a) We measured the semantic similarity of two MeSH terms in the MeSH taxonomy using Jiang similarity metric [8] according to the following formula:

$Sim(m_1, m_2) = 1 + \frac{ic(m_1)+ic(m_2)-2.ic(lc(m_1,m_2))}{2}$ wherein m_1 and m_2 are the MeSH terms, and $ic(m)$ is calculated according to the following method: $ic(m) = 1 - \frac{\log(|sc(m)|+1)}{\log(|M|)}$ where $sc(m)$ is the set of transitively subsumed (descendent) terms of m and |M| is the size (total number of terms) in M. $lc(m_1, m_2)$ is the "least common subsumer" (superclass) that subsumes both m_1 and m_2. b) We used researcher Matching algorithm [7] to find similarity between two practitioners using the following formula. Let u_1= $\{m_{1,1}, m_{1,2}m_{1,3}, \dots, m_{1,h}\}$ and u_2= $\{m_{2,1}, m_{2,2}m_{2,3}, \dots, m_{2,g}\}$ where $m_{i,j}$ is a MeSH term used by practitioner u_i. Hence, similarity between two users is defined as follows:

$fsim(u_1, u_2) = \frac{1}{|u_1|.|u_2|}\sum_{m1,i} sim(m_{1,i}, m_{2,j}).\theta(m_{1,i}, m_{2,j})$

To implement the framework, we transformed the discussion forum data to RDF using a Semantic Web programming language (Jena). We SPARQL (a Semantic Web query language) to retrieve and analyze the RDF discussion forum data.

3. Results and Discussion

Table 1 indicates the semantic similarity between the practitioners calculated based on the proposed approach in Section 3 along with the top five central practitioners (DC>380, CC>0.8) and the overlapping topics discussed by them. The topics were determined based on the MeSH terms noted in the practitioners' messages. Also, we found a total of 1,1179 MeSH terms derived from messages exchanged between practitioners of which 591 terms were overlapping.

Table1. Practitioners with high semantic similarity (A), top five central practitioners (B)

A: Semantic similarity analysis			B: Social network analysis				
User1	User2	Semantic similarity	Degree centrality (DC) {user, value}		Closeness centrality (CC) {user, value}		Overlapped topics among central users
P951	P952	0.6825	P951	812.965	P951	0.875	Homosexuality, Rectal Cancer, Surgery, Pancreas, Hernia obturator, Coronary artery bypass
P951	P959	0.3965	P952	552.952	P971	0.862	
P952	P989	0.3684	P971	522.953	P952	0.833	
P951	P968	0.3609	P959	410.975	P989	0.833	
P952	P971	0.3322	P989	383.999	P968	0.801	

Our analysis indicates that the centrality metrics, calculated with our proposed approach, are in line with the centrality measures produced by traditional SNA methods. Comparing the high-frequency MeSH terms discussed by central practitioners with the most-used terms by all the practitioners implies that nine out of top ten MeSH terms in both lists were the same. For example, 'Sentinel Lymph Node Biopsy' was a subject with the highest frequency among all the MeSH terms (2,752) of which 1,088 (around 40%) were utilized by central practitioners. This analysis assists in identifying the most important topics discussed in the discussion forum (see Table 1-B). The found MeSH terms used frequently by the central practitioners can lead us to identify the practitioners' interests, specialties, and so forth. Hence, we argue that our approach provides a more comprehensive method to identify influential practitioners which is based on (a) SNA based centrality measures and (b) semantics based content analysis. And, as a by-product we are also able to highlight the most popular topics being discussed within the discussion forum.

4. Conclusion

In our work, we present a semantic web-based approach to analyze a medical discussion forum to identify central practitioners. This is an alternate approach for analyzing medical discussions whereby we are able to incorporate medical terms in the social network analysis. This allows the identification of topics that were discussed by the central practitioners. Our approach can be extended to other medical social networks to identify key participants or topics based on an analysis of the discussion content.

References

[1] B. Hoser, A. Hotho, R. Jäschke, C. Schmitz, and G. Stumme, "Semantic Network Analysis of Ontologies," in *The Semantic Web: Research and Applications*, Y. Sure and J. Domingue, Eds. Springer Berlin Heidelberg, 2006, pp. 514–529.

[2] S. A. Stewart and S. S. R. Abidi, "Applying social network analysis to understand the knowledge sharing behaviour of practitioners in a clinical online discussion forum," *J. Med. Internet Res.*, vol. 14, no. 6, p. e170, 2012.

[3] T. Berners-Lee, J. Hendler, and O. Lassila, "The Semantic Web: A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities," *Scientific American*, pp. 29–37, May 2001.

[4] J. Golbeck and M. Rothstein, "Linking Social Networks on the Web with FOAF: A Semantic Web Case Study," in *Proceedings of the 23rd National Conference on Artificial Intelligence - Volume 2*, Chicago, Illinois, 2008, pp. 1138–1143.

[5] B. Padrón, M. Nogales, and A. Traveset, "Alternative approaches of transforming bimodal into unimodal mutualistic networks. The usefulness of preserving weighted information," *Basic and Applied Ecology*, vol. 12, no. 8, pp. 713–721, Dec. 2011.

[6] P. M. Newman and H. F. Durrant-Whyte, "Geometric projection filter: an efficient solution to the SLAM problem," 2001, vol. 4571, pp. 22–33.

[7] Y.-B. Kang, Y.-F. Li, and R. L. Coppel, "Capturing Researcher Expertise Through MeSH Classification," in *Proceedings of the 8th International Conference on Knowledge Capture*, New York, NY, USA, 2015, p. 6:1–6:8.

[8] T. Pedersen, S. V. S. Pakhomov, S. Patwardhan, and C. G. Chute, "Measures of semantic similarity and relatedness in the biomedical domain," *Journal of Biomedical Informatics*, vol. 40, no. 3, pp. 288–299, Jun. 2007.

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-491

Towards an Open Infrastructure for Relating Scholarly Assets

Christopher MUNRO[a,1], Philip COUCH[a], Jon JOHNSON[b], John AINSWORTH[a], and Iain BUCHAN[a]
[a] *Health eResearch Centre, Division of Informatics, Imaging, and Data Sciences, School of Health Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester Academic Health Science Centre, UK*
[b] *Centre for Longitudinal Studies, UCL Institute of Education, London, UK*

Abstract. Discovery of useful relationships between scholarly assets on the web is challenging, both in terms generating the right metadata around the assets, and in connecting all relevant digital entities in chain of provenance accessible to the whole community. This paper reports the development of a framework and tools enabling scholarly asset relationships to be expressed in a standard and open way, illustrated with use-cases of discovering new knowledge across cohort studies. The framework uses Research Objects for aggregation, distributed databases for storage, and distributed ledgers for provenance. Our proposal avoids management by a single central platform or organization, instead leveraging the use of existing resources and platforms across natural partnerships. Our proposed infrastructure will support a wide range of users from system administrators to researchers.

Keywords. Research informatics, distributed systems, publication archives, research objects, cohort studies, discovery networks, knowledge management.

1. Introduction

The tsunami of data generated, or leveraged by, social and biomedical sciences poses a significant challenge in knowledge discovery for both researchers and data investors (those collecting or enabling the collection of research-ready data).

For example, researchers may be interested in finding scholarly assets relevant to their work and describing how they have derived new knowledge from existing assets. At the same time, data investors may wish to track the outputs, such as published papers, based on their data – this is illustrated in a recent literature search by the Millennium Cohort Study wishing to track the outputs from its data [1].

The current models of knowledge discovery via structured metadata focus on large Data Archives and Research Data Portals. These centralized approaches involve the curation of metadata (often with extensive manual processes), primarily from retrospective documentation of surveys and other data collections. With the increasing availability of portals and platforms storing scholarly assets there is move to better reuse existing data to answer new questions [2]. To maximize the impact of this reuse asset

[1] Corresponding Author: Dr. Christopher Munro, Health eResearch Centre, Division of Informatics, Imaging, and Data Sciences, University of Manchester, Jean McFarlane Building, Oxford Road, Manchester, M13 9PL, UK. Email: chris.munro@manchester.ac.uk

producers need to specify the relationships between existing scholarly assets that asset consumers can navigate to discover the context of use.

To facilitate the discovery of new knowledge there needs to be open and shared specification of the aggregated assets, and an infrastructure to make these aggregations available. Such assets may be co-produced across multiple organizations/sources that do not necessarily share the source data, for example due to governance constraints. Akin to a car built of components from different factories.

The problem for discovery is two-fold: encouraging the creation of metadata around generated data resources and connecting metadata to its source through a chain of provenance at minimal additional cost. This discovery problem is like that of research using articles from serials and journals, which use exchange protocols such as OAI-PMH, and standards such as Z39.50. However, whereas the digital manuscript field benefits from having a small set of similar metadata models, in social and biomedical science data management there are many (and quite different) metadata models in use.

To address this problem we reviewed existing approaches, technologies, and standards, and we developed use-cases and specified requirements for an infrastructure to support discovery of relationships between scholarly assets.

2. Methods

2.1 Use Cases

We developed the use-cases below drawing on the experiences and needs of users of the HeRC e-Lab and CLOSER discovery platforms for collaborative research (across organizations and disciplines) using shared data sources.

The HeRC e-Lab is an online research collaboration platform which can combine and harmonize existing datasets, for example storing multiple birth cohort datasets, plus linked clinical data, as part of the STELAR project [3]**.** The CLOSER discovery platform [4] enables researchers to search and explore data from eight UK longitudinal studies.

We have considered a range of users from those driving research and consuming content from platforms, to the platform developers and data managers:

1. A researcher has obtained data from a Research Data Portal and generates some derived variables and wants to share what it is and how it was generated, and provide a citable link for publication.
2. A researcher is looking at a dataset available from a Research Data Portal and wants to know what other researchers have produced before deciding whether to explore a funding call.
3. A Principal Investigator wants to understand what data from their study has been used in published research and whether the data produced is being under-utilized.
4. A Platform owner would like to make their users' aggregations of assets searchable in a lightweight shared network, where there is no reliance on a single third party central platform being maintained and available.

2.2 Requirements

We identified the following requirements for a system to serve the use-cases above:

- Store assets in distributed infrastructures and search them in a unified way.

- Search and create aggregations of existing assets (e.g. publications, datasets), and specify relationships between them (via a web-based interface and programmatically for external platforms).
- Identify and authorize users, for example using OpenID, or ORCIC ID.
- Harvest assets' records and aggregations from existing platforms.
- Assist/guide users in creating 'profiles' for aggregations to ensure sufficient context is recorded and that the aggregations have enough metadata to be usefully searched.

3. Results

Cross platform/format search has been driven largely by serials, journals and libraries using the centralized approaches described below. The move toward common standards has been very helpful in this area, e.g. the CORE platform [5] which harvests multiple repositories via the OAI-PMH standard protocol, and journals using Z39.50.

Federated search: one platform facilitates search over all resources/service providers. The records from the service providers are stored at the data archive sites. The records' schema needs to be the same to easily add a new provider, unless a *mediator* is written to allow new record formats to be ingested, with additional cost. As data archives manage the assets the search interface is always up to date and there are fewer issues with synchronization. Search performance can be poor as the process involves waiting for the query results from each service provider.

Cross archive search via harvesting: data are first gathered from sources and then stored locally in the search platform, resulting in improved performance but with inherent synchronization issues. This is the approach of large search engines, but requires a common data format e.g. OAI-PMH. There is further complication if a source goes down and the record still exists in the search platform.

To enable discovery of content across platforms, we investigated a variety of possible approaches and technologies:

Aggregation*:* a key requirement for the infrastructure we propose is a method to aggregate resources that can be identified (e.g. dereferenced by a URI). The OAI-ORE standard for reuse and exchange defines Resource Maps [6] and these can also be made identifiable to allow discovery [7]. The Research Objects standard for aggregating resources [8] builds on OAI-ORE and incorporates formalized annotations, capture of provenance metadata, as well as minimum requirements for metadata, and the use of checklists [9] to assess the quality of aggregations.

Blockchain: there is emerging interest in utilizing Distributed Ledgers to benefit from the decentralized capabilities and provenance tracking properties of blockchain. The original implementation, bitcoin, is purely currency based but variants such as Multichain or Ethereum allow assets to be stored and blockchain additions validated. Blockchain has several weaknesses for this discovery role, e.g. it is not designed for large document storage, e.g. records of 1MB for Datacoin. The cost of adding to the block chain scales with the size of the block chain, this is challenging for an infrastructure designed to be used over a long period. Only the transaction information (but not the contents of the block chain) is searchable. Once records are added they are immutable which could problems in governance rich areas such as data access.

Distributed databases: distributed databases support the decentralized nature of our proposed infrastructure. Elastic search (which incorporates a search engine), and,

Couchbase are examples document stores and Cassandra is a columnar database. Graph variants include Titan, Neo4j and OrientDB with querying via the Gremlin graph querying language. Compared to blockchain these perform better for searching (in blockchain only the transaction history is searchable), and share the decentralized nature and redundancy but lack the in-built provenance capturing abilities. They also more naturally support file storage, for example Couchbase can store 'blob's of JSON up to 20MB. They also support the distributed nature (adding extra nodes), for example with Cassandra new nodes can be added to the cluster using a certificate approach.

3.1 Proposed Approach

Figure 1 contains an illustration of our approach. We propose an infrastructure that uses an aggregation format such as Research Objects (ROs), to be stored in a network of distributed databases, which supports the addition of new organizations. Identification and authorization could be handled via open ID, or ORCID ID [10] and usernames and passwords to cater for users unable to access the other methods.

The infrastructure incorporates a web interface for users (e.g. researchers, data investors) to add and search aggregations. The interface could be authenticated to, via Open ID or ORCID. This enables use of the system by those with low resources, whereas platform hosts may interact via software.

Both e-Lab and CLOSER are nodes in this example, and would host an instance of the distributed database, with the capability to push both records of content via OAI-PMH, and ROs.

The Figure illustrates the example of a dataset, contained in both the CLOSER portal and the HeRC e-Lab. There are ROs in the e-Lab aggregating existing data, scripts and results, and ROs in the CLOSER platform contain the questions, studies, sweeps and instruments.

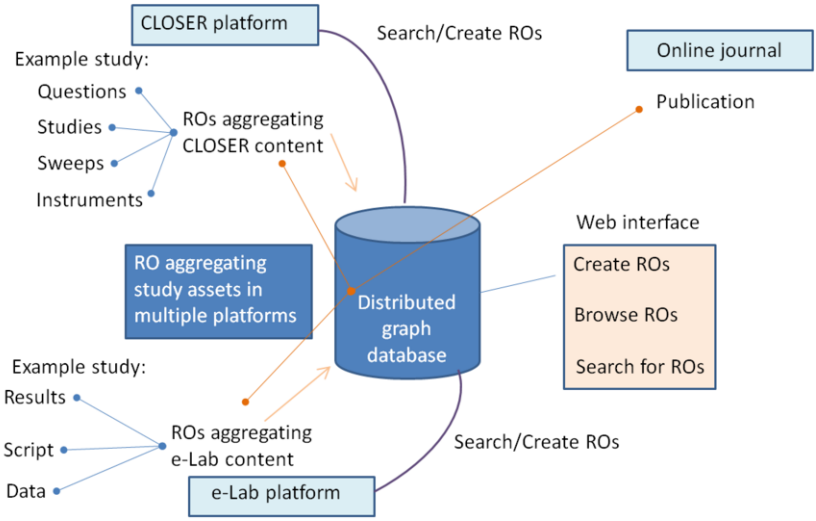


Figure 1: Outline of proposed asset relationship discovery framework, illustrated with e-Lab and CLOSER platforms and study data, using a currency of (extended) Research Objects (ROs).

4. Discussion

We have studied theoretical and practical solutions for discovering new knowledge from the relationships between digital scholarly assets – reviewing existing technologies and standards; exploring use-cases; specifying requirements; proposing an open technical and operational framework that can leverage existing platforms quickly.

Two key strengths of the proposed framework are: 1) it does not need to be maintained by a single organization – operations are distributed and thus resilient; and 2) it builds on existing standards, platforms, and assets.

The framework allows users (e.g. data investors, or researchers building on existing work) to aggregate related assets from different locations or platforms, enabling the discovery of relationships that were previously fragmented across the distributed assets. We propose using ROs to specify their aggregations and publish them in ways that are easy to find, share and reuse. The exchange of ROs provides a mechanism for discovery between different platforms – this can be a partial exchange where encapsulated source data may not be shared (e.g. resources are within a private portal, that requires additional access authorization) but their metadata are sufficient for discovery of new knowledge (e.g. contact details to request information from the portal). This highlights the need to specify profiles for 'types' of ROs, although this is currently missing from the RO specification, minimum information checklists [8] could be developed to facilitate different profiles.

The use of a distributed database for ROs increases availability by avoiding reliance on a single platform, and this approach enables trusted networks to be developed by incorporating extra nodes.

We are currently linking the e-Lab and CLOSER platforms using the framework presented here as part of UK Research Councils' move toward more collaborative research using shared digital assets.

References

[1] **D, Kneale, et al., et al.** *Piloting and producing a map of Millennium Cohort Study Data usage: Where are data underutilised and where is granularity lost?* London : EPPI-Centre, UCL Institute of Education, 2016.

[2] *Increasing value and reducing waste: addressing inaccessible research.* **Chan, Dr An-Wen, et al.,** 2014, The Lancet, pp. 257-266.

[3] *The Study Team for Early Life Asthma Research (STELAR) consortium 'Asthma e-lab': team science bringing data, methods and investigators together.* **Custovic, A, et al.,.** 2015, Thorax, pp. 799-801.

[4] *CLOSER Discovery.* [Online] https://discovery.closer.ac.uk/.

[5] CORE system. [Online] https://core.ac.uk/.

[6] Discovery of resource maps. [Online] http://www.openarchives.org/ore/1.0/discovery.

[7] Resource maps. [Online] http://www.openarchives.org/ore/1.0/primer.

[8] *Why Linked Data is Not Enough for Scientists.* **Bechhofer, Sean, et al.,** Future Generation Computer Systems, 2013, Vol. 29.

[9] *MIM: A Minimum Information Model vocabulary and framework for Scientific Linked Data.* **Gamble, Matthew, et al.,** 2012 IEEE 8th International Conference on E-Science (e-Science).

[10] Introduction to the ORCID Public API. *ORCID.* [Online] https://members.orcid.org/api/introduction-orcid-public-api.

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-496

Architecture and Initial Development of a Digital Library Platform for Computable Knowledge Objects for Health

Allen J FLYNN[a,b,1], Namita BAHULEKAR[a], Peter BOISVERT[a], Carl LAGOZE [b], George MENG[a], James RAMPTON[a], and Charles P FRIEDMAN[a,b,c]
[a] *Department of Learning Health Sciences of the Medical School,*
[b] *School of Information, and* [c] *School of Public Health at the University of Michigan*

Abstract. Throughout the world, biomedical knowledge is routinely generated and shared through primary and secondary scientific publications. However, there is too much latency between publication of knowledge and its routine use in practice. To address this latency, what is actionable in scientific publications can be encoded to make it computable. We have created a purpose-built digital library platform to hold, manage, and share actionable, computable knowledge for health called the Knowledge Grid Library. Here we present it with its system architecture.

Keywords. Knowledge Objects, Digital Library, Knowledge Management, Digital Preservation, Knowledge Sharing, Provenance

1. Introduction

This work addresses the latency between biomedical knowledge discovery and its use in practice [1]. It is founded on informatics methods to make biomedical knowledge computable [2, 3]. However, most biomedical knowledge that could be made machine-interpretable has not yet been [4, 5]. Also, the modest amount of computable biomedical knowledge that exists can be difficult to update, as it is often embedded within e-Health systems [6-8]. Further, computable knowledge about human factors or the information and implementation sciences also has the potential to improve eHealth applications [9-11]. The Knowledge Grid Library developed at the University of Michigan can store, preserve, manage, and disseminate reusable knowledge objects. Knowledge objects held in the library are comprised of binary code with metadata [12]. The library is part of a larger Learning Health System infrastructure [13]. This paper describes the system architecture of the Knowledge Grid Library, which has been realized after seven months of software development using Agile methods.

2. Research Question

The research question motivating this work is this: What are a minimum set of necessary and sufficient logical-technical components, which, when integrated, can bring about a

[1] Allen J. Flynn, Department of Learning Health Sciences of the Medical School, University of Michigan, 300 North Ingalls Building, 1161C, Ann Arbor, Michigan, USA, 48109; E-mail: ajflynn@umich.edu

stable, scalable digital library capable of **storing** knowledge objects, making them **discoverable and shareable**, facilitating their **incorporation** into various e-Health applications, and supporting object and library **management**?

3. Materials and Methods

To create an encompassing architectural overview of the Knowledge Grid Library, noting its logical-technical components, we analyzed the system architectures of several existing digital library platforms and other open source technologies. A logical-technical component is a single entity in the system architecture that can be implemented by various technical means.

The Fedora project brings a fundamental component to the Knowledge Grid Library's system architecture [14]. Fedora version 4.x is open source middleware that combines a native Resource Description Framework (RDF) linked data platform with binary file storage [15]. The Fedora repository middleware can scale up to store millions of holdings with their RDF descriptions. Two existing digital library platforms built on Fedora were also studied: Islandora and Hydra [16-17]. In addition to these, we studied the OpenICPSR resource, which is a Fedora-based data repository built using the Archonnex Architecture [18].

The California Digital Library provides a service called EZID to create and manage long-term globally unique identifiers (IDs) for digital objects [19]. EZID mints Archival Resource Keys (ARKs) that are used by the Knowledge Grid Library to solve the problem of uniquely identifying knowledge objects with universal identifiers. ARKs support finding these objects regardless of the deployed library instance where they are held.

Library holdings consist of digital knowledge objects. We have previously published a model of a digital knowledge object for health [12]. In technical terms, a knowledge object is software code (the payload), written in any programming language, that is annotated with RDF metadata (object description), linked to source evidence (object citations), linked to a license governing its use (object license), and also augmented with a RDF/XML representation of its inputs and outputs (object interface description). Examples of payloads include encoded predictive models, computable guidelines, and computable phenotypes.

4. Results

A library provides a set of core functions [20]. These core functions can be organized into three functional categories: **Storage** of holdings, **Discovery** of holdings, and **Management** of both holdings and the library itself. A successful digital library adequately addresses all three of these categories. A system architecture for a digital library of computable knowledge objects for health was designed with these three functional categories in mind (Figure 1).

What follows is a description of the Knowledge Grid Library's current capabilities to store, preserve, and protect computable knowledge objects for health. The capabilities described here have all been realized during the initial development phase of the library platform.

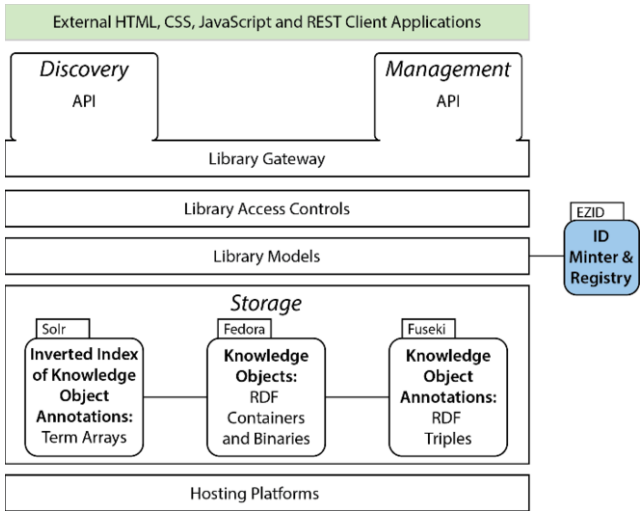


Figure 1. System Architecture of a Digital Library for Computable Knowledge Objects for Health

4.1. Storage

Knowledge objects held by the Library are stored in Fedora using a tree structure comprised of (a) RDF Containers, (b) Binary Containers, and (c) the branches between them. In Fedora, RDF Containers hold each object's metadata annotations and links to evidence and licenses while Binary Containers hold the executable software payloads of each object. Digital preservation is supported by Fedora's native fixity checking. It is possible to query Fedora directly through its Application Programming Interface (API). Integration of several storage sub-components is included in this architecture. RDF Triples are automatically extracted from Fedora and loaded into a separate Fuseki Triple Store. The Fuseki Triple Store enables SPARQL queries to be executed on the RDF object annotation data for all objects. An additional component which stores RDF annotation data is Apache Solr. Solr supports discovery by automatically generating an inverse index from RDF object annotation data.

4.2. Discovery

Besides its integrated Storage components, the Knowledge Grid Library's system architecture includes several other software components that we have built.

The first of these is the Library Models component, which is used to create knowledge objects according to predefined knowledge object models. The Library Models component also applies a specific provenance model – the W3C PROV-O ontology – to consistently author RDF statements that record key events in the lifecycle of knowledge objects [21]. A specific object metadata model is used by the Library Models component to enable users to create, read, update, and delete object annotations. Fields used for these annotations include several Dublin Core elements along with links to citations of published scientific works, and links to object licenses. An external system, EZID, is used by the Library Models component to request and receive unique Archival Resource Key (ARK) IDs for every knowledge object and to register the

whereabouts of each object on an external EZID server. EZID, working in conjunction with the Name-to-Thing Resolver, enables knowledge objects to be persistently located and accessed over the World Wide Web.

The second component we have built is the Library Access Controls component that is used to control access to the holdings stored in the Library.

We have also built a Library Gateway component with two APIs, one enabling external systems to do object discovery (Discovery API) and another to let external systems import, export, and update knowledge objects (Management API). The Library Gateway implements these two APIs.

4.3. Management

In the Knowledge Grid Library architecture, the Management API applies mainly to knowledge object management and less to overall Library management. The set of basic knowledge object Management capabilities currently supported by the Management API consist of mechanisms enabling external systems and applications to (1) create/upload, (2) "checkout" or read/download, (3) edit/update, and (4) delete knowledge objects from the library. In addition, the Library Models component includes encoded logic to govern all four of these object management capabilities, including user configurable options for setting up object version control. The knowledge object "checkout" capability provided by the Management API is particularly important because it enables external e-Health applications to automatically gain access to knowledge objects and apply the encoded knowledge they carry in their payloads to tasks that take place outside of the Knowledge Grid Library.

5. Conclusion

A digital library to hold health-related digital knowledge objects, which are comprised of annotated and curated software payloads that are "ready to run", differs from other digital libraries and code repositories. It provides secure storage for executable code and preserves and protects the software it holds over a long period of time. It makes the objects it holds readily discoverable, accessible, shareable, and usable and it enables client applications to create, update, "checkout", and apply the software payloads in knowledge objects. A system architecture for the Knowledge Grid Library that meets all of these requirements has been reviewed. It externalizes and modularizes computable health knowledge to support clinical decision support and faster knowledge dissemination within a Learning Health System.

Acknowledgements

The Knowledge Grid development team thanks the following individuals for their support: Tanner Caverly, Lisa Ferguson, Rocky Fischer, Oliver Gadabu, Zach Landis-Lewis, Bob Riddle, Wei Shi, M. Grace Trinidad, John Walsh, and Michele Wichorek.

References

[1] Z. S. Morris, S. Wooding, and J. Grant, "The answer is 17 years, what is the question: understanding time lags in translational research.," J Royal Soc. Med., vol. 104, no. 12, (2011), 510–520.
[2] R. N. Shiffman, G. Michel, A. Essaihi, and E. Thornquist, "Bridging the guideline implementation gap: a systematic, document-centered approach to guideline implementation," J Am Med Inf. Assoc, vol. 11, no. 5, (2004), 418–426.
[3] R. Studer, V. R. Benjamins, and D. Fensel, "Knowledge engineering: Principles and methods," Data and Knowledge Eng., vol. 25, no. 1–2, (1998), 161–197.
[4] C. Friedman, J. Rubin, J. Brown, M. Buntin, M. Corn, L. Etheredge, C. Gunter, M. Musen, R. Platt, W. Stead, K. Sullivan, and D. Van Houweling, "Toward a science of learning systems: a research agenda for the high-functioning Learning Health System," J. Am. Med. Informatics Assoc., (2014), 43–50.
[5] W. W. Stead, J. R. Searle, H. E. Fessler, J. W. Smith, and E. H. Shortliffe, "Biomedical Informatics: Changing What Physicians Need to Know and How They Learn," Acad. Med., vol. 86, no. 4, (2011), 429–434.
[6] R. S. Evans, S. L. Pestotnik, D. C. Classen, T. P. Clemmer, L. K. Weaver, J. F. Orme Jr., J. F. Lloyd, J. P. Burke, and J. F. Orme, "A computer-assisted management program for antibiotics and other antiinfective agents," N Engl J Med, vol. 338, no. 4, (1998) 232–238.
[7] P. J. Haug, B. H. S. C. Rocha, and R. S. Evans, "Decision support in medicine: lessons from the HELP system.," Int. J. Med. Inform., vol. 69, no. 2–3, (2003), 273–84.
[8] B. J. Ingui and M. A. Rogers, "Searching for clinical prediction rules in MEDLINE," J Am Med Inf. Assoc, vol. 8, no. 4, (2001) 391–397.
[9] M. W. Kreuter and V. J. Strecher, "Do tailored behavior change messages enhance the effectiveness of health risk appraisal? Results from a randomized trial," Heal. Educ Res, vol. 11, no. 1, (1996), 97–105.
[10] F. J. Seagull, "Human Factors Tools for Improving Simulation Activities in Continuing Medical Education," J. Contin. Med. Educ. Heal. Prof., vol. 32, no. 4, 2012.
[11] T. Madon, K. J. Hofman, L. Kupfer, and R. I. Glass, "Implementation Science," Science (80-.)., vol. 318, no. 5857, (2007), 1728–1729.
[12] A. Flynn, W. Shi, R. Fischer, and C. P. Friedman, "Digital Knowledge Objects and Digital Knowledge Object Clusters: Unit Holdings in a Learning Health System Knowledge Repository," Hawaii Int. Conf. Syst. Sci. 2016, 2016.
[13] C. Friedman and M. Rigby, "Conceptualising and creating a global learning health system," Int. J. Med. Inform., vol. 82, no. 4, pp. 1–9, 2013.
[14] C. Lagoze, S. Payette, E. Shin, and C. Wilper, "Fedora: an architechture for complex objects and their relationships," Int. J. Digit. Libr., vol. 6, no. 2, (2006) 124–138.
[15] Duraspace, "Fedora," (2015). Available: www.fedora-commons.org.
[16] Islandora Foundation, "Islandora," (2016). Available: www.islandora.ca.
[17] Project Hydra, (2016). Available: www.projecthydra.org
[18] T. Murphy and H. Ummerpillai, "The Architecture of Data Science and Archiving - Archonnex Architecture and Technology Stack," in IASSIST 2016, 2016.
[19] California Digital Library, "EZID", 2016. Available: www.ezid.cdlib.org
[20] Council on Library and Information Resource (CLIR), No brief candle: Reconceiving research libraries for the 21st century, August, 2008.
[21] W3C, "PROV-O The PROV Ontology." (2016) Available: https://www.w3.org/TR/prov-o/

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-501

An Approach for the Support of Semantic Workflows in Electronic Health Records

Marco SCHWEITZER[a,1] and Alexander HOERBST[a]
[a] *eHealth Research and Innovation Unit,*
UMIT - University for Health Sciences, Medical Informatics and Technology, Austria

Abstract. With the unprecedented increase of healthcare data, technologies need to be both, highly efficient for the meaningful utilization of accessible data and flexible to adapt to future challenges and individual preferences. The OntoHealth system makes use of semantic technologies to enable flexible and individual interaction with Electronic Health Records (EHR) for physicians. This is achieved by the execution of formally modelled clinical workflows and the composition of Semantic Web Services (SWS). Several seamless components provide a service-oriented structure defined by individual designed EHR-workflows. This work gives an overview of the planned architecture and its main components. The architecture constitutes the basis for the prototype implementation of all components. With its highly dynamic structure based on SWS, the architecture will be able to cope with both, the individual users' needs as well as the quick evolving healthcare domain.

Keywords. Electronic Health Records, Workflow, Diabetes Mellitus, Service-oriented architectures, Semantic Web Services

1. Introduction

A big challenge for stakeholders within the healthcare domain is to cope with the unprecedented amount of accessible healthcare data, e.g. clinical patient data collected through electronic health records (EHR). One of the major challenges is to make use of this data in a meaningful way in order to support efficiency and healthcare quality whilst keeping the health IT solutions flexible and adaptable to individual preferences and future changes [1]. Although research has advanced to address important challenges such as syntactic and semantic interoperability, major limitations on the functional level perceived by the EHR end users have not been tackled sufficiently.

Nowadays, semantic technologies play a key role in information utilization for many areas, even beyond the Semantic Web: The growing number of ontologies available for bio- and medical informatics purposes and the increasing number of knowledge-based systems for patient-related activities suggest potential benefits of these technologies [2].

The project OntoHealth is developing concepts for the flexible and individual execution of EHR-workflows using a semantic, service-based platform [3]. Physicians will be able to design and execute workflows, modeled as a set of tasks. Each task is basically defined by a certain functionality, inputs, outputs and non-functional properties. An individually created workflow is interpreted by the system as a template for a domain specific situation (e.g. "diabetes routine examination"). During the actual

[1] Marco Schweitzer, EWZ 1, 6060 Hall in Tyrol, Austria. Email: marco.schweitzer@umit.at

execution, suitable Web Services will be discovered and invocated by incorporating the workflow and context. The project aims to support EHR-usage with a new level of flexibility according to information processing in order to finally increase the usability of EHRs for healthcare professionals' daily work. This is especially relevant for interdisciplinary domains like diabetes Mellitus, where available information needs to be preprocessed according to the physician's needs and the patient's context. More information about the OntoHealth project can be found online[2]. In order to enable such a flexible and individual EHR information processing it is necessary to provide a suitable system architecture that fulfills the required goals. The present paper will explain the developed software architecture for OntoHealth as well as the methods followed to design the architecture.

2. Methods for architecture design

The OntoHealth approach empowers the physician to interact with the EHR by means of the execution of computerized patient-centered clinical workflows. These workflows refer to a set of activities conducted by the physician such as "get overview of the patient" or "prescribe medication" where interaction with medical record data is required. To derive informational, functional and non-functional requirements for the OntoHealth system we focused on Diabetes mellitus and conducted a systematic literature review [3], direct observations [4] and interviews with clinical personnel [5]. The results of these activities enabled us to identify common patterns and tasks within clinical workflows conducted in diabetes consultations, physician desired EHR-functions (e.g. "retrieve data", "visualize trends", "calculate risk score") regarding the utilization of needed medical patient data (e.g. "HbA1c", "blood pressure", "therapy", "referral") and non-functional parameters (e.g. "availability", "performance", "costs") users deem necessary when executing the workflows.

This information was used for the development of the semantic model that constitutes the backbone of the OntoHealth platform: The *standard-based ontology for clinical workflow-based e-health services in the domain of diabetes mellitus* (WISE-DM) [6] integrates six sub-models to semantically describe EHR-related clinical workflows, information needs, functions and non-functional properties. The same model is also used for the description of the semantic web services, which provide the functions and access to the EHR. The Minimal Service Model (MSM) [7], Business Process Modelling and Notation (BPMN) ontology [8] and the Web Service Modeling Ontology (WSMO) [9] were integrated and used for its development. WISE-DM contains two different types of EHR-workflow concepts: (1) workflow templates and (2) workflow routines. While a workflow template defines the sequence/structure of tasks and gathers requirements within each task for a common clinical situation (e.g. "diabetes routine examination"), the workflow routine is used for one concrete execution, i.e. for each task a list of concrete goals is created according to the user input of the physician and the given context, e.g. the related patient case and physician role.

Given the highly dynamic setting of the healthcare domain in general, it is necessary to provide flexible software which adheres to individual preferences and is able to cope with future challenges, e.g. new medical developments, organizational changes, legal restrictions. In order to incorporate all gathered requirements and the need for a dynamic

[2] www.ontohealth.org

structure, service-oriented architectures provide a suitable means [10]. Requirements were modeled using the Unified Modelling Language (UML), which resulted in use-case, activity and sequence diagrams. Those were used within several group discussions of the project participants in order to derive the architectural structure and its components. The architecture development process has followed an agile approach, where the architecture has been developed in recurring cycles of architecture modification and prototype implementation. Modifications are based upon the results of continuous evaluation according to the Deming-cycle principle [11].

3. Results: The OntoHealth architecture model

The OntoHealth architecture is based on the idea of service-oriented architectures. The conceptual view of the architecture (see Figure 1) describes the integration of the main *OntoHealth core*, the *Semantic Web Service Bus* and an existing *IHE XDS*[3] part.

The *Workflow Manager Platform* is in charge of the communication with the physician (or any other end user) through the *User Interface*, manages the workflows and consists of the *Workflow Engine*, which is responsible to create all semantic routine instances according to the template, user input and context. The *Workflow Loader* module enables to obtain the workflow templates available in the *Triple Store*. Beside the templates, the *Triple Store* will be aware of related routines, goals, the service descriptions, domain knowledge (e.g. data element class "HbA1c" with properties like "unit : mmHg") and data (data element instance, e.g. "blood pressure = 118/76 mmHg"). The *Workflow Engine* uses all information provided as user input and available in the *Triple Store*, i.e. the selected workflow template will be transformed to a workflow routine including context-aware goals. The goals are then executed according to the workflow structure by composing suitable semantic web services based on the selection results from the *Service Delivery Platform*.

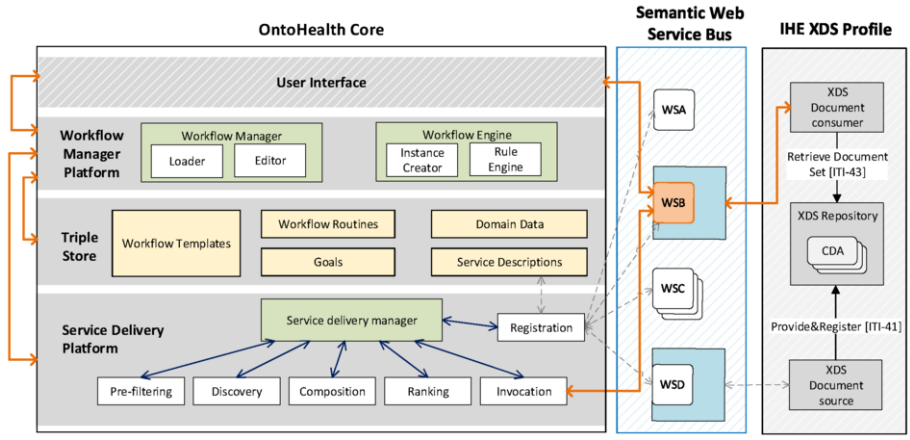


Figure 1. The OntoHealth architecture from a component-based view. The backend of the OntoHealth core constitutes of Workflow Manager Platform, Triple Store and Service Delivery Platform. The frontend with the user interface has not been specified in detail yet. Loosely coupled web services are provided through the Semantic Web Service Bus. Web services can interface an IHE-XDS based EHR. Orange arrows depict the relations between the components.

[3] Integrating the Healthcare Enterprise - Cross Document Sharing. See online: www.ihe.net

The service delivery process includes several steps ranging from registration of the services, pre-filtering, discovery (matching), composition of services, ranking, selection and finally invocation. As shown in architecture (see Figure 1), the *Service Delivery Platform* includes components to serve these functionalities, where the registration will be the first, resulting in a set of service descriptions within the triple store. A service description formally describes the capabilities of the web service consisting of inputs/outputs, function and non-functional properties.

Once a goal has been created for the workflow routine, the filtering process will discard all available services that do not match the mandatory non-functional requirements set by the physician from a functional point of view (e.g. "service should be available in Austria"). The aim of the discovery phase is to identify those services available in the *Semantic Web Service Bus* that fit a specific created goal and the requirements of the physician. The discovery of the services is done in two steps: (1) matching based on function classification and (2) matching based on inputs and outputs. Composition of services might be necessary if the available services do not fulfill all the requirements. Based on the user-selected, non-functional properties, the ranking module will return an ordered, ranked list of services, which are fulfilling the requirements best.

The final phase in a service delivery process is the actual invocation of the service, which was previously discovered, ranked and selected. Invocation enables the actual service to be called and the results to be delivered. Consuming the functionality of a service might happen in a straightforward manner, e.g. by simply invoking one service operation and returning the result, or a more complex interaction with the service is required in order to get its results. This particular aspect of the service interaction, dealing with how to invoke and interact with a service is known as service choreography [12]. It is often the case that more complex interaction is required in order to consume the service functionality, e.g. calling several service operations in a certain order, the order of the calls depends on the output of the previous operation, etc. As part of the invocation component, the OntoHealth *Service Delivery Platform* includes a service choreography engine, which is responsible for "understanding" the service choreography and guiding the interaction with the service in order to get its results. One service may have access to the EHR (e.g. based on IHE-XDS) to obtain the adequate clinical information while others may have functionality to extract data from a given source.

Results of a certain web service may be either used for (1) further consumption of services in terms of using it in a composition or for the subsequent task within the workflow or (2) to provide them for the *User Interface Module*, which will then be responsible to display the requested information according to the user's needs.

4. Discussion and Conclusions

Semantic technologies play an important role in our approach, enabling firstly to formally describe functions, data elements, and non-functional properties involved in a clinical workflow and secondly facilitating the selection of the most suitable and available eHealth services that fulfill these requirements.

The OntoHealth architecture presented in this paper includes some innovational aspects, which distinguishes the design from other works: (1) User requirements are managed through formal models of EHR workflows that might be individually designed. Including also non-functional properties opens new possibilities in user-centered EHR-execution. (2) A workflow engine is capable of executing the workflows involving the

context of a certain situation (e.g. the patient-case). This can be used to adapt a workflow (template) for one patient-case, e.g. when additional information is needed, or data should be processed in a different way. (3) The description of eHealth web services is also modeled in a formal manner based on the WISE-DM ontology. This ontology was developed very thoroughly by incorporating all necessary stakeholders within the development process. Appropriate reasoning will be used to discover, compose and rank suitable services or service-compositions. (4) Flexible Invocation through service choreography deals with the communication and potentially side effects (e.g. mediation, error handling). The architecture provides dynamic interfaces required for communication with IHE-related components.

Agile development with continuous evaluation has proven so far to be a suitable means for the present prototype development as design results can be checked immediately in implementation. However, the actual capability of this method will be evaluated when the prototype is finalized.

The next steps within the project are to continue the prototype implementation for all components and finalize the service-oriented architecture design. Related web services will further be developed in order to enable concrete executions for test cases.

Acknowledgments

This work has been supported by the Austrian Science Foundation (FWF), project number P 25895-B24.

References

[1] T.H. Payne, S. Corley, T.A. Cullen, T.K. Gandhi, L. Harrington, G.J. Kuperman, et al., Report of the AMIA EHR-2020 Task Force on the status and future direction of EHRs, *J. Am. Med. Informatics Assoc.* **5** (22) (2015), 1102–1110.

[2] N.F. Noy, M. Dorf, M.J. Montegut, N.H. Shah, N. Griffith, D.L. Rubin, et al., BioPortal: A Web repository for biomedical ontologies and data resources, *CEUR Workshop Proc.* (401) (2008), 1–4. http://protege.stanford.edu/conference/2009/abstracts/S3P1Noy.pdf.

[3] M. Schweitzer, N. Lasierra, S. Oberbichler, I. Toma, A. Fensel, A. Hoerbst, Structuring Clinical Workflows for Diabetes Care-An Overview of the OntoHealth Approach., *Appl Clin Inf.* **5** (2) (2014), 512–26.

[4] M. Schweitzer, N. Lasierra, A. Hoerbst, Observing health professionals' workflow patterns for diabetes care - First steps towards an ontology for EHR services., *Stud. Health Technol. Inform.* (210) (2015), 25–29.

[5] M. Schweitzer, N. Lasierra, A. Hoerbst, Requirements for workflow-based EHR systems - Results of a qualitative study, *eHealth 2016* (under revi) (2016), 124–131.

[6] N. Lasierra, M. Schweitzer, T. Gorfer, I. Toma, A. Hoerbst, Building a semantic model to enhance the user's perceived functionality of the EHR, *Stud. Health Technol. Inform.* (228) (2016), 137–41.

[7] T.V. Jacek Kopecký, Karthik Gomadam, Minimal Service Model, (n.d.),. http://iserve.kmi.open.ac.uk/ns/msm/msm-2014-09-03.html (accessed February 11, 2016).

[8] M. Rospocher, C. Ghidini, L. Serafini, An ontology for the business process modelling notation, *Front. Artif. Intell. Appl.* (267) (2014), 133–146.

[9] World Wide Web Consortium (W3C), Web Service Modeling Ontology (WSMO), (n.d.),. https://www.w3.org/Submission/WSMO/ (accessed May 1, 2016).

[10] E. Vasilescu, S.K. Mun, Service Oriented Architecture (SOA) implications for large scale distributed health care enterprises, in: Conf. Proc. - 1st Transdiscipl. Conf. Distrib. Diagnosis Home Heal. D2H2 2006, 2006: pp. 91–94.

[11] R. Moen, C. Norman, Evolution of the PDCA Cycle, *Society* (2009), 1–11.

[12] D. Roman, M. Kifer, Semantic Web Service choreography: Contracting and enactment, in: Lect. Notes Comput. Sci., 2008: pp. 550–566.

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-506

Identifying Emerging Trends in Medical Informatics: A Synthesis Approach

Yasmin VAN KASTEREN[a], Patricia A.H. WILLIAMS[b] and Anthony MAEDER[a,1]
[a] *School of Health Sciences, Flinders University, Adelaide SA, Australia*
[b] *School of Computer Science, Engineering & Mathematics, Flinders University, Adelaide SA, Australia*

Abstract. Medical informatics is a young and rapidly evolving field, influenced by and impacting on many different knowledge domains. Recent contributions on scoping the associated body of knowledge are confounded both by variations in popular use of terminology for established areas, and by the advent of new areas without yet established terminology. Determining the scope of a topic through online bibliographic search filters is a well-established approach in scientific research and has been developed as a human-directed task. Establishing the best approach and automating the process has proved a difficult problem. This paper explores the use of text analysis of bibliographic information using available search engines and NVIVO text analysis tools to test the potential for dynamic word based filters based on data mining. Results show that word searches of abstracts are more effective than topic searches for identifying health informatics papers, however more work is required to refine search terms to improve generalisability. Using data mining to track changes in word use in medical informatics journals, may make it possible to establish a more dynamic search filter to match the evolving nature of the field of health informatics.

Keywords. medical informatics, knowledge discovery, search engine, text analysis

1. Introduction

Since the late 1990s, the field of Medical Informatics has evolved. The term Medical Informatics is used in this research as a generic term rather than exclusive one. Terms used by other authors include biomedical informatics, clinical informatics, health informatics, health information technology, as well as more popular terms including e-health, digital health, digital medicine. Medical Informatics is still a comparatively young field, with many of its intrinsic concepts still being refined, and with many new directions proposed. It is influenced by, and in turn impacts upon, different knowledge domains such as health, technology, social sciences, and humanities. The rapidity of change in an age of health systems reform, and the impact of disruptive influences such as those of new technologies or of swinging political directions, make it desirable to track emerging trends in the field. This would allow incremental adjustment of the acknowledged scope of the field to incorporate these extensions, which in turn will assist in establishing currency in education, research and policy development for the field.

[1] Email: anthony.maeder@flinders.edu.au

2. Scope of Medical Informatics

Recent work which was undertaken has adopted a structured and detailed approach to provide a scoping of the field. IMIA [3] developed a knowledge-based list of 48 topic areas with specific recognition of two major domains: (a) medicine, health and biosciences, health system organization; (b) informatics and computer science, mathematics, biometry. AMIA [4] took a competencies-based approach and identified five broad categories for these: (a) fundamental scientific skills; (b) scope and breadth of the discipline; (c) theory and methodology; (d) technological approach; (e) human and social context. These approaches offer a sound basis for maintaining and extending descriptions of the body of knowledge for the field, as they offer fixed hierarchical structures into which new concepts can be inserted and classified.

Nevertheless, there are limitations in such a structural approach, which become evident when elements need to be added. As demonstrated Elkin et al [5] mapping between complementary structures may be necessary, to provide a holistic view of the scope of the field, whilst a fuller interpretation of elements in the structures may not conform to that mapping. Martin-Sanchez and Lopez-Campos [6] identified another limitation by showing how adding elements associated with some new area of knowledge may intersect with existing structures in numerous ways [e.g. 6].

Determining the scope of a topic through online bibliographic search filters is a well-established approach in scientific research and has been developed as a human-directed task (e.g. in the undertaking of systematic reviews of the literature [7]). Establishing the best approach and automating the process has proved a difficult problem, to which some current attention is being directed [8]. This paper documents the development and testing of search filters based on common word usage in bibliographic information and explores ways in which the selection of words used in word based search filters might be automated to enable dynamic search filters that remain current by tracking change in word use over time.

3. Methodology

A three-step method was used to construct and test a word based search filter for medical informatics based on text analytics of common word usage in bibliographic information.

Identifying key search words and the optimal search field. Individual word frequency searches using a stemmed word search (e.g. health, healthy, healthful) was performed using NVIVO, on bibliographic information (abstract, title, keywords) of four leading journals in the field of medical informatics (2001-2015): Journal of Medical Internet Research (JMIR), Methods of Information in Medicine (MIM), Journal of the American Medical Informatics Association (JAMIA) and International Journal of Medical Informatics (IJMI). The journals were selected for their diversity of coverage based on European and US vocabulary, and applied and theoretical papers covering both medical and informatics journal. To capture the diversity of terminology a preliminary search developed and tested search terms for each journal. The top 10% of words in each of the resulting word lists, ordered by word count, were then manually coded into one of three categories: health, informatics or neither. These three categories are used as the research is restricted to either health/medical based words, informatics related or neither of these. Subsequently, a series of Scopus searches was performed to test the sensitivity of the top

ten words for each journal (health and informatics) using each journal as its own reference standard. (Sensitivity is measured as a percentage by dividing the number of records retrieved by the total number of relevant records). Searches were also conducted to identify which bibliographic field (title, abstract, keyword) resulted in the highest sensitivity.

Comparing key word searches with topic word searches. After removal of duplicates, all the top ten search word stems both informatics and health from all four journals from Step 1 were combined as follows:

> *adopt* OR algorithm* OR app OR application* OR apps OR code* OR coding OR comput* OR data OR EHR* OR electronic* OR imag* OR informatics OR Internet OR mobil* OR online OR predict* OR record* OR system* OR System* OR technolog* OR tool* OR user* OR web*) AND (biomedical or cancer* or care* or caring or clinic* or disease* or drug* or health* or heart* or hospital* or medic* or nurs* or patient* or physician* or prescribe* or treatment* or weigh* or well*

The results of the word based above were compared to a broad topic Scopus search which included terms such as digital/e-health, health/medical/clinical informatics, health information technology and electronic/digital health/patient records. Having identified that abstracts produced the highest matched results for word searches, a comparison of the performance of the word versus topic base searches was undertaken on all four journals. These were limited to abstract searches only.

Exploring word use changes over time. To explore how word use changes over time might be tracked semi-automatically, abstracts from one of the four journals (JMIR) were separated into three five-year time periods. NVIVO, stemmed word frequency searches were used to identify the top 5% of most frequently used words. To separate the informatics and health words from other commonly used words, the top 5% of words were semi-automatically coded using Excel's vlookup function. The vlookup table array was based on all the results of the manual coding of informatics and health words in Step 1 above. Results for each time period were ranked based on word frequency because the number of papers published was variable across the years: 2001-2005 (n=183) 2006-2010 (n=251) 2011-2015 (n=1437).

4. Results

Results from Step 1 (Table 1) show that abstracts are the best for sourcing and for matching word based searches with a sensitivity of between 80% and 98% compared to between 48% and 67% for word searches based on titles and keywords. Results also showed that there was variance in word use between the four journals. The word *technology* was in the top ten of all four journals. *Comput*, data* and *system** were only top ten in three out of four journals. *Web** was only prominent in JMIR and system* popular in all but MIM. There was more concurrence in the top ten health words terms. The search terms *health*, patient*, medic*, clinic** and *physician** were commonly used across all four journals. Other health search words only appeared in one journal for example *biomedical* (MIM) and *drugs* (JAMIA).

Table 1. Results of word searches using abstract, title and keyword for the journal JMIR and MIM

	JMIR (n=1889)		MIM (n=966)	
Health words	**Retrieved**	**Sensitivity**	**Retrieved**	**Sensitivity**
Abstracts	1743	92%	774	80%
Titles	1125	60%	477	49%
Keywords	1186	63%	463	48%
Informatics words				
Abstracts	1842	98%	886	92%
Titles	1036	49%	477	49%
Keywords	1272	67%	611	63%

The results of Step 2 comparing word based search with topic based searches are shown in Table 2. Word searches in each of the four medical informatics journals showed results of between 86% and 100%, in each case higher than the equivalent topic search which only resulted in specificities of between 31% and 70%.

Table 2. Results comparing topics search and word search

	JMIR (n=1871)		MIM (n=852)		JAMIA (n=1770)		IJMI (n=1561)	
	Ret*	Spe^	Ret*	Spe^	Ret*	Spe^	Ret*	Spe^
Word based search TITLE-ABS-KEY	1833	98%	839	98%	1767	100%	1341	86%
Not captured in search	38	2%	13	2%	3	0%	220	14%
Topics based search TITLE-ABS-KEY	573	31%	346	41%	1235	70%	916	59%
Not captured in search	1298	69%	506	59%	535	30%	644	41%

* Retrieved ^ Specificity

Finally, word use (both health and informatics) changes over time. Results show changes in the top ten ranked words as described in Step 3. Health is consistently the highest ranked word (based on word frequency) over all three time periods. Other word rankings changed over time with words like medical and cancer and community being replaced with the words programs, data and systems. There were also changes in the lower rankings with the words appearing in 2011-2015 included: *app, twitter, smartphone* and *mhealth*, respectively ranked 16, 45, 46 and 54. These additions reflect changes in trends and language use, with the use of the abbreviated apps rather than applications, lexical changes with the introduction of the word mhealth (mobile health), changes in social trends with the inclusion of twitter and lastly changes in technology with the inclusion of smartphone.

5. Limitations

The four journals chosen for the research are not necessarily representative of the field and a different choice of journals might result in different search terms. Scopus was used for convenience but other search engines such as Medline or Google Scholar may very well produce different results. Equally the topic search terms used in Step 2 were generated by the third named researcher. Different topics choices may result in different outcomes. The word searches tested in this paper need to be further refined to improve generalisability. It is likely that some of the search terms are too generic to be used in abstract searches in other journals. Combining words (e.g. *electronic* and *record*) and/or

replacing existing words with words outside the top ten, more specific to the field of medical informatics, might produce more generalisable without too much loss of specificity.

6. Discussion & Conclusion

This paper offers an early contribution towards the development of a systematic process for identifying emerging trends in the field, by mining of recent literature using available search engines and text analysis tools and offers suggestions on ways in which the selection of search words can be updated so filters remain dynamic.

A number of findings emerge. First, word based searches retrieve more papers than do topic based searches, when tested against journals in the field of medical informatics. When using data mining to identify words to use in a word based search, mining needs to be conducted across a number of journals as word use and frequency of word uses varies across journals. Word based searches retrieve the most papers when searches are conducted using abstracts rather than title or keywords. Abstracts have more consistent and natural word use than do titles and keywords and also contain more information.

Second, word use changes over time, not only the relative frequency of word use but also the emergence of new words. New word use emerges for multiple reasons including changes in everyday word use, social trends and technology. One of the biggest challenges still to be addressed is automating the identification of health and informatics words. In this paper vlookup was used to automate the coding of words, however this approach requires a reference source or lexicon. An alternative solution would be to create a comprehensive stop word list to eliminate words that are related to neither health or informatics. In conclusion, text mining to identify key search words and change in word use over time can facilitate and potentially automate the creation of dynamic filters to capture the cross disciplinary and evolving nature of the field of medical informatics.

References

[1] Bernstam, E. V., Smith, J. W., & Johnson, T. R. (2010). What is biomedical informatics?. *Journal of Biomedical Informatics*, 43(1), 104-110.

[2] Haux, R. (2010). Medical informatics: past present and future. *International Journal of Medical Informatics*, 79, 599-610.

[3] Mantas, J., Ammenwerth, E., Demiris, G., Hasman, A., Haux, R., Hersh, W., ... & Wright, G. (2010). Recommendations of the International Medical Informatics Association (IMIA) on education in biomedical and health informatics-first revision. *Acta Informatica Medica*, 18(1), 4.

[4] Kulikowski, C. A., Shortliffe, E. H., Currie, L. M., Elkin, P. L., Hunter, L. E., Johnson, T. R., ... & Smith, J. W. (2012). AMIA Board white article: definition of biomedical informatics and specification of core competencies for graduate education in the discipline. *Journal of the American Medical Informatics Association*, 19(6), 931-938.

[5] Elkin, P. L., Brown, S. H. and Wright, G..(2013). Biomedical informatics: we are what we publish. *Methods of Information in Medicine*, 52, 538-546.

[6] Martin-Sanchez, F.J. and Lopez-Campos, G.H. (2016). The New Role of Biomedical Informatics in the Age of Digital Medicine. *Methods of Information in Medicine*, 55, 392-402.

[7] Lefebvre, C., Glanville, J., Wieland, L. S., Coles, B., & Weightman, A. L. (2013). Methodological developments in searching for studies for systematic reviews: past, present and future?. *Systematic Reviews*, 2(1), 1.

[8] Cohen, A. M., Smalheiser, N. R., McDonagh, M. S., Yu, C., Adams, C. F., Davis, J. M. and Yu, P.S. (2015). Automated confidence ranked classification of randomized controlled trial articles: an aid to evidence-based medicine. *Journal of the American Medical Informatics Association*, 22, 707–717.

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-511

Appraising Healthcare Delivery Provision: A Framework to Model Business Processes

Daniela LUZI[1], Fabrizio PECORARO and Oscar TAMBURIS
National Research Council. Institute for Research on Population and Social Policies, Rome, Italy

Abstract. Children are dependent on a reliable healthcare system, especially for the delivery of care which crosses the primary/secondary care boundary. A methodology based on UML has been developed to capture and single out meaningful parts of the child healthcare pathways in order to facilitate comparison among 30 EU countries within the MOCHA project. A first application of this methodology has been reported considering asthma management as an example.

Keywords. Business process modelling, UML, rich picture, building blocks, children, asthma

1. Introduction

The main objective of the MOCHA (Models of Child Health Appraised) project[2] is to compare and appraise existing models of primary care for children among 30 EU/EEA countries. The complexity and variation in care provision makes it a challenging task that needs a holistic and multidisciplinary approach for the analysis of health service patterns, and measurement of system performance, outcomes and costs. The identification of optimal models of primary child healthcare considers how health systems organise the interaction and integration of primary, secondary and social care to deal with the different healthcare needs, ranging from well-being related services, that monitor the psychophysical development of a healthy child and/or of children with special conditions or needs, to the treatment of acute and chronic conditions. Under this perspective the application of business process modelling contributes to the analysis of the current delivery models adopted in each EU country providing a synthetic and pictorial description of the different ways of organizing, coordinating and delivering children's care. For this reason, to produce a consistent and scientifically sound business process of child healthcare delivery a conceptual framework was developed to capture and single out meaningful parts of the child healthcare pathways so to facilitate the comparison between national health systems.

The paper is structured as follows: a methodology that captures the business process from a high-level to a low-level of granularity is presented in Section 2, providing the description of each model applied as well as its use in the MOCHA perspective. The third section reports the application of the methodology using asthma disease management as an example. Finally, discussion and conclusions are presented.

[1] Daniela Luzi, IRPPS CNR, via Palestro, 32 – 00185 – Rome, Italy; E-mail: d.luzi@irpps.cnr.it
[2] MOCHA website available at http://www.childhealthservicemodels.eu/

2. Methods

The methodology proposed in this paper is described in Table 1, highlighting the main steps to be followed to identify homogeneous parts of the process of child care that can facilitate the comparison between national health systems.

Table 1. Main steps of the proposed methodology highlighting aim, method, input and result of each step.

Aim	*Method*	*Input*	*Result of the step*
Conceptual model			
Identify the main components that influence the child care from the MOCHA perspective	Rich picture	MOCHA objectives; Vision of child care	Determinants of care, set of macro-processes.
Identify the macro-processes and macro-activities that describe the child care	Map of building blocks	Rich picture	Identification, for each macro-process, of the generic workflow and of the linked building blocks.
UML model			
Identify the activities, the actors involved and their relationships to a disease	Use case diagrams	Macro-processes; Guidelines	Identification, for each macro-process, of all possible actors and roles involved.
Identify similarities and differences between countries in terms of: Activities performed; Actors involved; Where it is executed; Timeline	Activity diagrams	Use cases; Case studies; National norms and regulations	Identification, for each use case, of a set of activity diagrams grouping countries with similar process performance. They identify activities and messages exchanged, actors involved, their relationships and interactions.

The highest level of abstraction is described by a *rich picture* [1,2] that specifies the main components setting the boundaries of the healthcare service provision centred on child care pathways.

The second point of view, called *map of building blocks*, is a refinement of the rich picture that specifies the macro-processes describing their generic workflows and related macro-activities to be analysed in the provision of child care. Each map that represents one of the main healthcare pathways is composed by a sequence of building blocks executed in parallel or in series to accomplish a well-defined objective. Decision points are also included in the workflow to determine alternatives as well as parallel paths within a process flow. Additional linked building blocks are included in the map and represent macro-activities that can be executed in certain circumstances and anytime within the main process. Differently from the use of building blocks in software development [3], in our approach they are needed to single out homogeneous, comparable parts of the process that have to be taken into account when analysing the different national care provision.

This conceptual point of view is the starting point to apply and develop the UML models [4] that provide a standard way to visualize the process description relevant for our aims. In our methodology we select the use case and the activity diagrams as suitable means to describe the interaction between the different stakeholders as well as the activities performed in each identified scenario [5]. The *UML use case diagram* describes a single, previously identified building block focusing on a specific disease or type of intervention, generally based on available clinical guidelines. Considering that clinical guidelines are commonly agreed by the scientific community, they can be used as an ideal set of activities to be performed, thus providing a country-independent picture as an important point of reference to assess the quality of care provision.

Finally, once the actors and macro-activities of a specific building block have been identified, *UML activity diagrams* are used to capture the activities performed and the

messages exchanged by the different actors as well as triggering conditions taking also into account the location and timeline of each activity. The description of the activity diagrams is based on the results of Country Agents questionnaires developed by MOCHA partners around case studies, that capture real-life context [6] in child health service provision. On the basis of these descriptions, countries that have similar procedures, use similar services, and are based on similar caregivers for the provision of care are grouped and then compared.

3. Results: Application of the Methodology

This paragraph gives an example based on asthma treatment that illustrates how UML use case and activity diagrams (UML Model from Table 1) have been integrated in the proposed methodology. The rich picture shown in Figure 1a models at the centre of the system the child represented as a girl jumping from well-being to disease management and vice versa. Children are generally a healthy population whose psychophysical development is monitored especially, but not only, in the first years of their lives. When an illness occurs, this is treated following a disease management pathway whose results lead to the normal well-being monitoring, or with the management of chronic conditions. Strictly connected with these two pathways social support is included in the rich picture to identify services addressed to improve quality of life, such as social inclusion, ability to live independently, fundamental rights. Moreover, these pathways are surrounded by the family and the related living and working conditions, the community as well as health services and professional caregivers. Other determinants of health are identified in socio-economic and financial conditions as well as cultural and environmental settings.

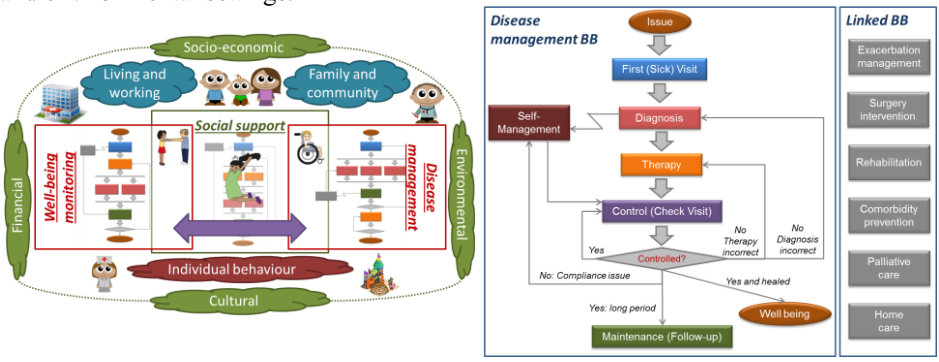


Figure 1. (a) Rich picture; (b) Disease management map of building blocks

Figure 1b shows a map of building block as a first refinement of the rich picture focusing on the disease management pathway. A high-level workflow comprises the macro-activities performed to make a diagnosis and prescribe the related initial therapy. A check visit is then scheduled to verify the appropriateness of the therapy, to control whether the issue is solved (return to well-being pathway) or follow-up activities should be foreseen (management of chronic patients). The disease management process comprises various linked building blocks that can represent the connection between primary and secondary care and/or the treatment of children with complex needs, such as exacerbation management, surgery intervention, rehabilitation. Figure 2 depicts a use case diagram that describes the “first sick visit” building block structured on the

GINA (Global Initiative for Asthma) [7] guidelines. Best practices described in guidelines do not specify who performs the activities given that it depends on the way the healthcare system is organized. In our perspective, this aspect is crucial to compare the different models of service provision in the 30 EU countries. Therefore, the use case diagram has the important function to identify actors and roles in the performance of the "first sick visit" macro-activities. This represents a synthetic, not country-specific, view useful to capture, for example, whether a diagnostic test is performed directly by a gatekeeper or a specialist and, in the first case, whether this is done by a GP or a Paediatrician depending on the national healthcare system (e.g. Italy is a primary care paediatrician-based system, whereas the UK is a generalist GP-based system).

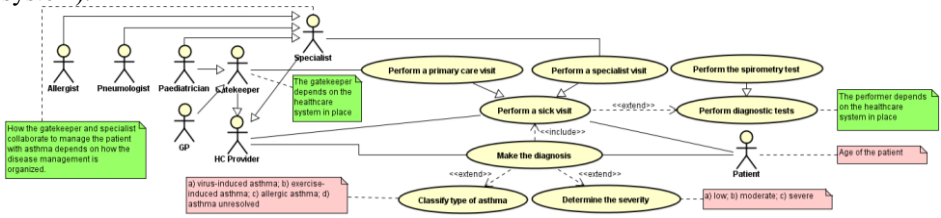


Figure 2. UML use case diagram describing a sick visit performed for the treatment of asthma

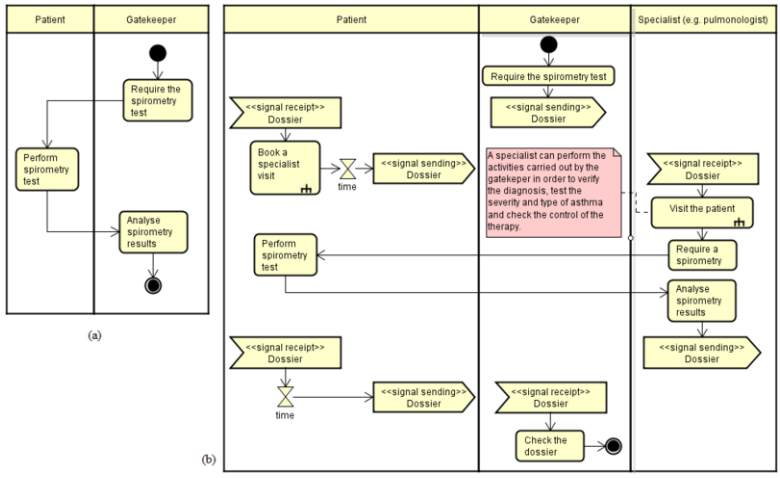


Figure 3. Activity diagram describing the spirometry test performed by a) gatekeeper; b) specialist

Figure 3 shows the activity diagram for the use case: "perform the spirometry test". The diagram is based on a case study where a ten-years-old child has occasionally experienced wheezy attacks and needs to perform this test during the sick visit to help the healthcare providers to make the diagnosis as well as determine the type and the severity of asthma. This test is an important part of the diagnosis process as it can be performed directly by the gatekeeper (GP or paediatrician) or by a specialist. Therefore, we can distinguish the countries in these two groups and compare the activities performed in each one. In particular, in Figure 3a the test is directly executed by the gatekeeper during the first sick visit. In this case the gatekeeper can analyse the results and consequently make the diagnosis. Differently, in Figure 3b the spirometry is performed by a specialist, for instance in countries where the gatekeeper is not equipped with spirometer or is not qualified to perform this task. Thus, in the diagram the introduction of this third actor is needed and visualized using the relevant swim

lane. In particular, the gatekeeper requires the family of the child to book a specialist visit during which the required test is performed. This implies the collaboration between the two healthcare professionals as well as the inclusion of administrative procedures related to the booking and eventually payment of the specialist visit. The two processes are compared not only on the basis of activities performed and actors involved but also on the time required to perform the diagnosis. This implies also to include waiting time for a child to access the primary as well as the secondary care visit, as shown by the hourglass reported in the diagram 3b.

4. Conclusions and Discussion

The paper describes a framework to facilitate the comparison between the different business processes performed in 30 EU/EEA countries in the delivery of child care within the MOCHA project. Given the difficulties in the a-priori identification of pre-defined models of care, the proposed methodology is based on a conceptual framework that allows the comparison between a set of homogeneous building blocks that are identified in the context of three main pathways: well-being monitoring, disease management and social support. The methodology also specifies how UML use case and activity diagrams are applied to achieve the comparison of business processes across countries. In particular, the use case diagram provides a snapshot of actors and activities related to a building block and relies on clinical guidelines that represent a benchmark to compare healthcare business processes. Activity diagrams highlight differences and similarities in activities performed, actors involved as well as timelines in real-life context. This methodology has the advantage of making comparison of child healthcare systems on the basis of homogenous, well-defined parts of the process of child care. Moreover, it is flexible enough to allow the analysis of parts of the process selected on the basis of the case studies identified by the MOCHA partners for the investigation and appraisal of health systems. Finally, this methodology is also suited to consider the interaction between primary and secondary care and/or complex care as well as the analysis of interfaces with social services within national healthcare systems.

Acknowledgment

This study is part of the Horizon 2020 MOCHA project. Grant agreement no. 634201.

References

[1] P. Checkland, J. Scholes, *Soft Systems Methodology in Action*, John Wiley and Sons, Chichester, 1990.
[2] H. Liyanage, D. Luzi, S. de Lusignan, F. Pecoraro, R. McNulty, O. Tamburis, P. Krause, M. Rigby, M. Blair, Accessible Modelling of Complexity in Health (AMoCH) and associated data flows: asthma as an exemplar, *Journal of Innovation in Health Informatics* **23** (2016), 476-484.
[3] J.K. Müller, Aspect design with the building block method, *Software Architecture* (1999), 585-601.
[4] J. Rumbaugh, I. Jacobson, G. Booch, *Unified Modeling Language User Guide,* Addison-Wesley, 2005.
[5] H. Eriksson, P. Magnus, *Business modeling with UML, Business Patterns at Work*, John Wiley & Sons, New York, USA, 2000.
[6] S. Crowe, K. Cresswell, A. Robertson, G. Huby, A. Avery, A. Sheikh, The case study approach, *BMC Medical Research Methodology*, 100 (2011), 1-9.
[7] GINA (Global Initiative for Asthma) guidelines. Available at www.ginasthma.org

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-516

A Case Study on Sepsis Using PubMed and Deep Learning for Ontology Learning

Mercedes ARGUELLO CASTELEIRO[a], Diego MASEDA FERNANDEZ[b], George DEMETRIOU[a], Warren READ[a], Maria Jesus FERNANDEZ PRIETO[c], Julio DES DIZ[d], Goran NENADIC[a,e], John KEANE[a,e], and Robert STEVENS[a,1]
[a]*School of Computer Science, University of Manchester (UK)*
[b]*Midcheshire Hospital Foundation Trust, NHS England (UK)*
[c]*Salford Languages, University of Salford (UK)*
[d]*Hospital do Salnés de Villagarcia, SERGAS (Spain)*
[e]*Manchester Institute of Biotechnology, University of Manchester (UK)*

Abstract. We investigate the application of distributional semantics models for facilitating unsupervised extraction of biomedical terms from unannotated corpora. Term extraction is used as the first step of an ontology learning process that aims to (semi-)automatic annotation of biomedical concepts and relations from more than 300K PubMed titles and abstracts. We experimented with both traditional distributional semantics methods such as Latent Semantic Analysis (LSA) and Latent Dirichlet Allocation (LDA) as well as the neural language models CBOW and Skip-gram from Deep Learning. The evaluation conducted concentrates on sepsis, a major life-threatening condition, and shows that Deep Learning models outperform LSA and LDA with much higher precision.

Keywords. Ontology Learning, Deep Learning, PubMed, OWL, SPARQL

1. Introduction

Sepsis is defined as a life-threatening organ dysfunction caused by a dysregulated host response to a new infection [1]. *Severe sepsis* occurs when complicated with organ dysfunction, and may progress to *septic shock* [1]. *Severe sepsis* and *septic shock* affect millions of people worldwide, and their incidence is increasing [2]. According to NHS England: "*sepsis now claims more lives than lung cancer, the second biggest cause of death after cardiovascular disease*" [3]. Severe sepsis is regarded as "*a common, deadly, and costly complication in cancer patients*" [4].

The Surviving Sepsis Campaign (SSC), established in 2002, updates consensus definitions and evidence-based guidelines for management of severe sepsis and septic shock. Evidence from the literature improves understanding of sepsis, leading to better patient care. However, the size and rate of growth of PubMed – the largest biomedical resource – is a challenge for term extraction and knowledge representation.

Ontology learning from text aims to "*turn facts and patterns from an ever growing body of information into shareable high-level constructs*" [5]. This paper investigates how the traditional distributional semantics methods, Latent Semantic Analysis (LSA)

[1] Corresponding author, E-mail: Robert.Stevens@manchester.ac.uk

[6] and Latent Dirichlet Allocation (LDA) [7], as well as the neural language models CBOW (Continuous Bag-of-Words) and Skip-gram of Mikolov et al. [8] from Deep Learning can facilitate ontology learning from an unannotated biomedical corpus.

2. Methods

In distributional semantic models (DSMs), the semantics of terms within a domain are determined empirically [9]. LSA is spatially motivated, while LDA is a probabilistic method. LSA and LDA have high computational and storage cost associated with building or modifying the model. However, CBOW and Skip-gram from Deep Learning make it feasible to obtain word embeddings (i.e. distributed word representations) from corpora of billions of words. Using similarity measures (e.g. cosine value for CBOW and Skip-gram) we can obtain the n top-ranked candidate terms for a query term (e.g. sepsis). Hence we can quantify empirically how closely related are two terms from a DSM.

This study takes advantage of the UMLS (Unified Medical Language System) Metathesaurus from the U.S. National Library of Medicine that contains more than three million biomedical concepts. Each UMLS concept has a unique identifier (a.k.a. CUI). UMLS Metathesaurus concepts are grouped into UMLS Semantic Types, where semantic types can also be merged into semantic groups [10].

We use OWL-DL [11] to formally represent concept names, concept expressions, and terminological axioms. We started by creating programmatically a small ontology that contains mostly the UMLS Semantic Types and Groups. This ontology together with lemon (Lexicon Model for Ontologies) [12] is key to the move from *candidate terms* from DSMs to UMLS Metathesaurus concepts with known synonyms and relationships. In this study, the concept *Lexicon* from lemon represents a vocabulary for a DSM, while the concept *Lexical entry* represents a single term (one or more words) in the vocabulary/lexicon of the DSM. We also link the concept *Lexical sense* from lemon with the UMLS Metathesaurus concept, an OWL class we created that can have as a subclass any of the more than 3 million UMLS Metathesaurus concepts.

A vocabulary/lexicon from DSMs contains lexical entries that are: concepts, phraseological expressions (typically combination of concepts), or spurious terms without a true biomedical meaning. UMLS MetaMap [13] can indicate which terms from the lexicon are UMLS Metathesaurus concepts and their UMLS Semantic Type(s).

Experimental setup – We downloaded the MEDLINE/PubMed baseline files for 2015 and also the update files up to 8th June 2016. Applying the PubMed Systematic Reviews filter [14], a subset of 301,202 PubMed publications (title and abstract) with date of publication from 2000 to 2016 was obtained. We performed two experiments using LSA, LDA, CBOW and Skip-gram: *Experiment I* – the pre-processing performed preserves capitalisation and numbers in the text; and *Experiment II* – after the pre-processing that preserves capitalisation and numbers in the text, Part-Of-Speech (POS) tagging and chunking was performed. Chunking labels segments of a sentence with syntactic constituents – e.g. noun phrase (NP) and verb phrase (VP). For the experiments reported here, we used gensim [15] and word2vec [16].

Domain expert evaluation – three medical consultants (rater A, B, and C) assessed the relevance of the terms in pairs (query term and candidate term) using a Likert-type (categorical) scale taken from [17]. According to this scale, a candidate term can be: not at all relevant (marked as 0); a little relevant (marked as 1); quite a bit

relevant (marked as 2); and very much relevant (marked as 3). Simple guidelines were given to the domain experts. They consist of: a) the Likert-type (categorical) scale; and b) a few examples illustrating pairs of query term-candidate term annotated with different scores. The inter-annotator agreement is calculated using two well-known measures: weighted Cohen's Kappa and Fleiss. Using few raters and simplistic annotation guidelines, low inter-annotator agreement and some difficulties are expected before arriving at consistent figures [18].

Task-based evaluation – the automatically extracted ontology – called here PubMed Ontology LEarning Ontology for Sepsis (POLEOS) – refers to the OWL-DL ontology built programmatically out of the n top-ranked candidate terms obtained from each model (i.e. LSA, LDA, CBOW, and Skip-gram) in *Experiment I* and *II*. POLEOS reuses the UMLS Semantic Types and Groups formalised in OWL. Two annotation properties were introduced in POLEOS to capture: a) the relevance of a candidate term, which comes from a voting system that considers equally the assessment made by each human rater per candidate term; and b) to what extent the candidate term was recognised by UMLS MetaMap. Using conventional measures in information retrieval [5], the suitability of the different models to find relevant terms (concepts or phraseological expressions) for sepsis is assessed. The numbers needed to calculate precision are obtained by executing SPARQL [19] queries over POLEOS. We can also query POLEOS with SPARQL to determine: the number of unique UMLS concepts needed to provide the biomedical meaning for the candidate terms; and how many candidate terms can be assigned to each UMLS Semantic type or group based on the *Lexical sense* assigned.

3. Results

Computational resources and execution times – The DSMs are generated using a Supermicro with 256GB RAM and two CPUs Intel Xeon E5-2630 v4 at 2.20GHz. The time to obtain a DSM goes from less than 1 hour for CBOW (the quickest) to more than 23 hours for LDA (the slowest). Using a MacBook Pro Retina (2.8 GHz Intel Core i7 with 16GB RAM) and Jena ARQ [20] as the SPARQL query engine, the mean time for executing a SPARQL query over POLEOS three times was less than 2 seconds.

Distributional Semantics: LSA, LDA, CBOW, and Skip-gram –The query term was "sepsis" for all models. For LDA in *Experiment II*, two topics for sepsis were identified, and thus, LDA in *Experiment II* has almost double the candidate terms than the other models.

Domain expert evaluation – Rater C did not score any candidate term with 0 (not at all relevant); while rater A scored 52 of the candidate terms with 0. Based on this finding, we decided to merge the four scores into two categories: *more relevant* with score 3 or 2; and *less relevant* with score 1 or 0. These two categories allow arrival at consistent figures, where the weighted Cohen's Kappa between two consultants is: 0.53 for rater A and B; 0.61 for rater A and C; and 0.55 for rater B and C. Fleiss Kappa is 0.56 for the three raters. In Figure 1, each column corresponds to a model with its name at the top. A column has up to three colours: dark grey for *most relevant* (score 3) candidate terms agreed by all three raters, grey for *more relevant* (score 3 or 2) agreed by at least two raters, and white *less relevant* (score 1 or 0) candidate terms agreed by at least two raters.

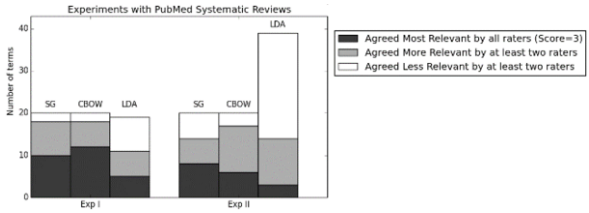


Figure 1. Agreed assessment by medical consultants for candidate terms, where SG = Skip-gram

Task-based evaluation – POLEOS contains a total of 3276 axioms, and its Description Logic expressivity is ALEHI(D). From the two experiments and the different models, a total of 157 candidate terms related to sepsis were obtained. The same candidate term may appear in more than one model; 130 candidate terms are unique: 71 are UMLS concepts, 50 are phraseological expressions (typically a combination of UMLS concepts), and 9 were unrecognised by UMLS MetaMap. Only 7 of the 9 unrecognised are truly spurious terms (i.e. terms that do not have a true biomedical meaning); the other two (i.e. suPAR and IgGAM) are acronyms that could be mapped to UMLS concepts. A total of 169 unique UMLS concepts are represented as subclasses of the concept *Lexical sense* from lemon. In other words, 169 UMLS concepts are needed to provide the biomedical meaning of the 157 candidate terms obtained. Three UMLS Semantic groups stand out with more candidate terms (concepts or phraseological expressions): *Disorders* (DISO) with 61 candidate terms, *Concepts & Ideas* (CONC) with 47 candidate terms, and *Living Beings* (LIVB) with 24 candidate terms. Table 1 shows the precision for each model in *Experiment I* and *II*.

Table 1. Task-based precision calculated as tp/(tp+fp) per model and experiment, where tp stands for the number of true positives agreed by at least two raters and fp stands for the number of false positives agreed by at least two raters.

Model	Experiment I	Experiment II
LSA	-	26%
LDA	58%	36%
CBOW	**90%**	**85%**
Skip-gram	**90%**	70%

4. Discussion and Conclusion

Healthcare data is multi-source, high volume and multi-modal. Identifying patterns in such data requires scalability while accommodating structured and unstructured data. Unlike conventional datasets, healthcare data is often incomplete and noisy; in turn, unlike standard analytics, neural language models process raw natural language data to associate terms with vectors of real valued features and place semantically related terms close together in the vector space [21]. The learned "*high level*" semantic features of the word embeddings are usually not explicitly present in input such as biomedical literature.

Hu et al. [22] recently reported improvements in the word embeddings generated by introducing POS tagging information into a neural network similar to CBOW. However, from the precision obtain for Experiment II (see Table 1) it is difficult to derive a real benefit from the chunking (VP and NP) performed.

Although several parameter settings were tried, it is plausible that other options may have improved the precision for LSA and LDA. However, a higher performance for CBOW and Skip-gram when compared with traditional DSMs such as LDA is aligned with the results reported in the literature for other studies, e.g. [23].

Using sepsis as the query term, we have shown how to anchor plain text candidate terms from DSMs into shareable high-level constructs, i.e. UMLS concepts represented in OWL that can be easily queried with SPARQL based on Semantic types and Semantic Groups. The evaluation performed indicates high precision for the neural language models, which in turn hints at the plausible acquisition of relevant medical concepts for sepsis. Hence, this paper illustrates how CBOW and Skip-gram can be used to aid ontology learning tasks for sepsis, a major healthcare problem, using unannotated PubMed Systematic Reviews citations (titles and abstracts) as a corpus.

Acknowledgements

This work was supported by grant 603288 from the European Union Seventh Framework Programme (FP7/2007-2013) for sysVASC project.

References

[1] Singer, M., et al. 2016. The third international consensus definitions for sepsis and septic shock (sepsis-3). Jama 315 (2016), 801-810.
[2] Levy, M.M., et al. Surviving Sepsis Campaign: association between performance metrics and outcomes in a 7.5-year study. Intensive care medicine 40 (2014), 1623-1633.
[3] NHS Report, https://www.england.nhs.uk/wp-content/uploads/2015/08/Sepsis-Action-Plan-23.12.15-v1.pdf Accessed Feb 2017.
[4] Williams, M.D., et al.,. Hospitalized cancer patients with severe sepsis: analysis of incidence, mortality, and associated costs of care. Critical Care, 8 (2004).
[5] Wong, W., Liu, W. and Bennamoun, M. Ontology learning from text: A look back and into the future. ACM Computing Surveys (CSUR) 44 (2012).
[6] Landauer, T.K. and Dumais, S.T. A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. Psychological review 104 (1997).
[7] Blei, D.M., Ng, A.Y. and Jordan, M.I. Latent dirichlet allocation. Journal of machine Learning research 3 (2003), 993-1022.
[8] Mikolov, T., Chen, K., Corrado, G., and Dean, J. Efficient estimation of word representations in vector space. arXiv pre-print arXiv:1301.3781 (2013).
[9] Cohen, T. and Widdows, D. Empirical distributional semantics: methods and biomedical applications. Journal of biomedical informatics 42 (2009), 390-405.
[10] UMLS Semantic Network, https://semanticnetwork.nlm.nih.gov Accessed Feb 2017.
[11] OWL 2 Web Ontology Language, https://www.w3.org/TR/owl2-overview/ Accessed Feb 2017.
[12] lemon - The Lexicon Model for Ontologies, http://lemon-model.net Accessed Feb 2017.
[13] MetaMap, https://metamap.nlm.nih.gov Accessed Feb 2017.
[14] PubMed SB, https://www.nlm.nih.gov/bsd/pubmed_subsets/sysreviews_strategy.html Accessed Feb 2017.
[15] Gensim, https://radimrehurek.com/gensim/ Accessed Feb 2017.
[16] word2vec, https://code.google.com/archive/p/word2vec/ Accessed Feb 2017.
[17] Aaronson, N.K. Quality of life assessment in clinical trials: methodologic issues. Controlled Clinical Trials 10 (1989), 195-208.
[18] Biemann, C. Ontology learning from text: A survey of methods. In LDV forum 20 (2005), 75-93.
[19] SPARQL 1.1 query language, https://www.w3.org/TR/sparql11-query/ Accessed Feb 2017.
[20] Jena ARQ, http://jena.sourceforge.net/ARQ/ Accessed Feb 2017.
[21] LeCun, Y., Bengio, Y. and Hinton, G. Deep learning. Nature, 521 (2015), 436-444.
[22] Hu, B., Tang, B., Chen, Q. and Kang, L. A novel word embedding learning model using the dissociation between nouns and verbs. Neurocomputing, 171 (2016), 1108-1117.
[23] Liu, Y., Liu, Z., Chua, T.S. and Sun, M. Topical Word Embeddings. In AAAI (2015), 2418-2424.

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-521

Acquisition of Expert/Non-Expert Vocabulary from Reformulations

Edwige ANTOINE[a] and Natalia GRABAR[a1]
[a] *Univ. Lille, CNRS, UMR 8163 - STL - Savoirs Textes Langage, F-59000 Lille, France*

Abstract. Technical medical terms are complicated to be correctly understood by non-experts. Vocabulary, associating technical terms with layman expressions, can help in increasing the readability of technical texts and their understanding. The purpose of our work is to build this kind of vocabulary. We propose to exploit the notion of reformulation following two methods: extraction of abbreviations and of reformulations with specific markers. The segments associated thanks to these methods are aligned with medical terminologies. Our results allow to cover over 9,000 medical terms and show precision of extractions between 0.24 and 0.98. The results and analyzed and compared with the existing work.

Keywords. Reformulation, Information Extraction, Medical Terminology, Layman Language, Patient Education

1. Introduction

Experts from medical area use sophisticated technical terms, which are usually non-understandable by patients [1]. This makes patients unable to make decisions, understand consequences of disorders and treatments, or even understand their disease. The situation is not improved by information increasingly available online [2]: patients remain often unable to understand it either. Find and understand new information implies solitary information retrieval process [3] when the patient is not accompanied by his medical doctor, with whom he may have verbal interactions and obtained needed information [4,5]. Indeed, with verbal interactions, it is possible to share knowledge and create common basis through reformulations, repetitions and clarifications [2].

Our purpose is to acquire vocabulary which associates specialized terms with layman expressions. We propose to exploit reformulations in texts written by experts or issued from collaborative media, in order to guarantee a better reliability of the extracted data. According to our hypothesis, (1) when experts reformulate terms, this indicates that the term is technical and conveys specialized meaning; (2) the reformulation act may allow to associate term with its reformulation; (3) reformulation is language phenomena spontaneously used in different kinds of texts.

[1] Corresponding author, Univ. Lille, CNRS, UMR 8163 - STL - Savoirs Textes Langage, F-59000 Lille, France. natalia.grabar@univ-lille3.fr

In what follows, we present some existing work (section 2). We describe the material and methods used (section 3), and then present and discuss the results (section 4). We conclude with some directions for future work (section 5).

2. Related Work

Reformulation occurs when someone is saying or writing again a given idea with different words [6], often in order to improve the understanding. Reformulation can be introduced by specific markers (*eg*, *let's say, That is*). Reformulated segments are not always semantically equivalent [7], but when they are it becomes possible to extract the paraphrases of technical terms. We can distinguish two main directions of the existing work: health literacy and automatic acquisition of paraphrases for medical terms. Health literacy is related to the understanding of medical information, its use and interpretation, and depends on two factors: environment (education, age, medical history...) and knowledge of patients, be it shared or not with the medical staff [8]. Several works have been dedicated to the acquisition of vocabularies associating terms with their paraphrases: exploitation of monolingual comparable corpora, in with morpho-syntactic patterns, similarity measures or n-grams allow to associate syntactic groups from both corpora [9,10]; exploitation of monolingual corpus for the acquisition of paraphrases for neoclassical compounds (*myocardial, desmorrhexia...*) thanks to the morphological analysis and segmentation of such terms, translation of their components in French (*myocardial: myo=muscle, cardia=heart*), and search of syntactic groups that contain these words (*heart muscle, muscle of the heart*) [11]. Let's also mention (1) the Consumer Health Vocabulary (CHV) in English [12], which a collaborative initiative, involving corpora processing, associative measures, and human annotators. The resource contains currently 141,213 unique terms; (2) an automatic translator of medical terms [13], best known through the resource MEDLINEplus, organized according to medical topics (tumors, obesity, etc). Let's also mention work on extraction of paraphrases, mainly applied to parallel and aligned corpora [14].

3. Material and Methods

We use the French part of the UMLS [15], and SNOMED International [16] (323,964 terms). We also use two corpora: (1) development corpus, issued from *masante.net* forum moderated by medical doctors. When users ask questions they are answered by the moderators. We use part with answers containing 6,139 answers (315,362 occ.); and (2) test corpus built with articles from the medical part of Wikipedia, which gives 18,434 articles (15,235,219 occ.). Finally, we use linguistic resources: 111 stopwords and morphological resources with 163,823 wordpairs (*eg* {*aorta, aortic*}) [17].

Our method is composed of four steps: (1) *pre-processing of corpora* with the syntactic analyzer Cordial [18]. Table 1 presents an example for *Vous devez les faire brûler par un gastroentérologue spécialisé, c'est-à-dire un proctologue (You must to make them burn by specialized gastroenterologist, that is a proctologist)*. The exploited

fields are forms, lemmas, and syntactic information *type GS* and *Prop*; (2) *extraction of reformulations*, which is done with two approaches:

- for *abbreviations*, we use the raw corpus and extract two structures: *extended form (abbreviation)* and *abbreviation (extended form)*. For this, we implement an existing algorithm, which allows to associate each letter from abbreviation with a given word before of between parentheses [19]. Three situations are possible: *full* when all letters from abbreviation are associated, *partial* when part of letters are associated, *null* when no letters are associated;
- for *reformulations with markers*, we exploit three markers *c'est-à-dire (That is), autrement dit (in other words), encore appelé (also called)* in the structure *concept marker reformulation*, like in *specialzed gastroenterologist, that is a proctologist*, where the underlined segments correspond to the source and target segments. It appears that source segments are better extracted with the *type GS* information, while target segments with the *Prop* information.

Table 1. An excerpt from syntactically tagged and analyzed text

form	lemma	POS	POSMT	GS	Type GS	Prop
Vous	vous	PPER2P	Pp2.pn	1	S	1
devez	devoir	VINDP2P	Vmip2p	2	V	1
les	le	PPER3P	Pp3.pa	3	C	2
faire	faire	VINF	Vmn--	4	D	2
brûler	brûler	VINF	Vmn--	5	V	3
par	par	PREP	Sp	8	F	3
un	un	DETIMS	Da-ms-i	8	F	3
gastroentérologue	gastroentérologue	NCMS	Ncms	8	F	3
spécialisé	spécialisé	ADJMS	Afpms	8	F	3
,	,	PCTFAIB	Ypw	-	-	3
c'	ce	PDS	Pd-..-	11	N	3
est	est	ADV	Rgp	-	p	3
-à	à	PREP	Sp	14	I	3
-dire	dire	VINF	Vmn--	14	I	3
un	un	DETIMS	Da-ms-i	8	F	3
proctologue	proctologue	NCMS	Ncms	16	D	3
.	.	PCTFORTE	Yps	-	-	-

(3) *alignment of the extracted segments with medical terminologies* has double objective: check the relevance of extractions and associate the extracted segments with medical terms. During the alignment, the extracted segments are normalized (accents, morphologically-related words), the stopwords are removed; and (4) *evaluation*, for which we build reference sets with two independent annotators for annotating source and target segments. For the evaluation of alignment with terminologies, we build a reference set from the development corpus. On the basis of the reference annotations, we can evaluate precision P, recall R and F-measure F of the extractions and alignments. Evaluation of extractions is performed with exact (boundaries of segments must be respected) and inexact (boundaries of segments can be inexact) versions.

4. Results and Discussion

The inter-annotator agreement [20] of extractions at the word level is 0.967 and 0.816, for alignments it is 0.208 and 0.714 for abbreviations and markers, respectively.

In the upper part of Table 2, we indicate number of extractions for each method: with abbreviations and the test corpus, we extract several candidates. With abbreviations, we observe three cases: correct extraction *{ESF; Editions Sociales Françaises}, {CD26; cluster de différenciation 26}*; partial correct extraction *{CHUM; Université Montréal}, {CHU; hôpital universitaire}*; partial incorrect extraction *{SEPP; plus}, {NFS; faire sang}*. Extractions with markers provide few duplicates because reformulations are less controlled *(des canaux galactophores: qui fabriquent le lait de la femme, qui sécrètent le lait*). Evaluation of extractions indicates that abbreviations show 0.74 and 0.94 F-measure, while markers show 0.24 and 0.98 F-measure with exact (borders respected) and inexact versions, respectively. In the lower part of Table 2, we indicate number of alignments. Within aligned segments, we can observe 5 cases: full correct *{syndrome polyalgique idiopathique diffus; syndrome polyalgique idiopathique diffus.C0016053}*; morpho-syntactic variation *{troubles fonctionnels intestinaux; troubles gastrointestinaux fonctionnels/C0559031}*; partial *{semaines amenorrhee; amenorrhee/C0002453}*; compositional *cause/C0085978* and *pus/C0034161* for *cause de pus*; and incorrect (*{LCR; ph lcr/C0853364}*, *{liquide cerebro; regime liquide/C-F2300}*). The average F-measure for the two segments is 0.71 and 0.73 with abbreviations and markers, respectively.

Table 2. Number of extractions and alignments for each method

	Development corpus		Test corpus	
	Abbrev.	**Markers**	**Abbrev.**	**Markers**
Extraction: nb occurrences	75	96	88,762	2,710
Extraction: nb types	42	96	8,106	2,710
Alignment: nb occurrences	75	96	88,762	2,757
Alignment: full alignments	11	5	154	42
Alignment: partial alignments	44	37	1,634	557
Alignment: not aligned	20	54	6,318	1,937

Table 3. Comparison with methods from the existing works

Method	Type of terms	Nb. extractions	Precision
Abbreviations	abbreviations	42, 8,106	0.74/0.94
Markers	any type	96, 2,710	0.24/0.98
Definitions [11]	any type	1,028	0.52, 0.68
Morphology [11]	compounds	1,128	0.76, 0.86
N-grams [9]	morpho-syntactic	65, 82	0.67, 0.60
Syntactic groups [10]	morpho-syntactic	109	0.66
Abbreviations [18]	abbreviations	785	0.95

In Table 3, we propose a comparison with existing works: type of terms, number of extractions, precision (available for all cited works). We can see that our methods are efficient: they provide an important number of extractions with good precision.

5. Conclusion and Future Work

For the acquisition of vocabulary associating technical terms with layman expressions, we propose to exploit reformulation through two methods: extraction of abbreviations and their extended forms, and of reformulations introduced by markers. The methods are fixed on the development corpus and then applied to the test corpus. Exact precision is between 0.23 and 0.74, while inexact precision is between 0.68 and 0.98.

The future work may study other markers. With the annotated corpora, we may apply supervised machine learning for making the extractions. The acquired vocabulary will be used for the simplification of medical and health documents.

Acknowledgement

This work has been performed as part of the *EQU (Éthique Qualité Urgence)* project funded by the call Thématique de l'établissement of université Lille 3.

References

[1] McCray A. Promoting health literacy. *J of Am Med Infor Ass* 2005;12:152-63.
[2] Zeng Q and Tse T. Exploring and developing consumer health vocabularies. *JAMIA* 2006;13:24-9.
[3] Boubé N and Tricot A. *Qu'est-ce-que rechercher de l'information ?: état de l'art*. Presses ENSSIB 2010.
[4] Vergely P, Condamines A, Fabre C, et al. Analyse linguistique des interactions patient/médecin. *Actes éducatifs de soins* 2009;92(5).
[5] Zielstorff RD. Controlled vocabularies for consumer health. *Journal of Biomedical Informatics* 2003;36(4-5):326-3.
[6] Le Bot MC, Schuwer M, and Richard E (dir.). *La reformulation : Marqueurs linguistiques – Stratégies énonciatives*. Rivages linguistiques, Rennes, 2008.
[7] Gulich E and Kotschi T. Les marqueurs de la reformulation paraphrastique. *Cahiers de linguistique française* 1983;5:305-51.
[8] Zeng QT, Kim E, Crowell J, and Tse T. A text corpora-based estimation of the familiarity of health terminology. In: ISBMDA 2006, 184-92.
[9] Deléger L and Zweigenbaum P. Paraphrase acquisition from comparable medical corpora of specialized and lay texts. In: AMIA 2008, 146-50.
[10] Cartoni B and Deléger L. Découverte de patrons paraphrastiques en corpus comparable: une approche basée sur les n-grammes. In: TALN, 2011.
[11] Grabar N and Hamon T. Extraction automatique de paraphrases grand public pour les termes médicaux. In: TALN 2015, Caen, France. 14 p.
[12] Zeng QT, Tse T, Divita G, et al. Exploring lexical forms: first-generation consumer health vocabularies. In: AMIA 2006, 1155
[13] McCray A, Loane R, Browne A, and Bangalore A. Terminology issues in user access to web-based medical information. In: AMIA Symposium 1999, 107-7.
[14] Androutsopoulos I and Malakasiotis P. A Survey of Paraphrasing and Textual Entailment Methods. *Journal of Artificial Intelligence Research,* 2010:38, 135-187
[15] Lindberg D, Humphreys B, and McCray A. The unified medical language system. *Methods Inf Med* 1993;32(4):281-91.
[16] Côté RA, Rothwell DJ, Palotay JL, Beckett RS, and Brochu L. *The Systematised Nomenclature of Human and Veterinary Medicine: SNOMED International*. College of American Pathologists, Northfield, 1993.
[17] Grabar N and Zweigenbaum P. A general method for sifting linguistic knowledge from structured terminologies. *JAMIASUP* 2000:310-4.
[18] Laurent D, Nègre S, and Séguéla P. Apport des cooccurrences à la correction et à l'analyse syntaxique. In: TALN, 2009.
[19] Schwartz AS and Hearst MA. A simple algorithm for identifying abbreviation definitions in biomedical text. In: Pacific Symposium on Biocomputing, 2003:451-6.
[20] Cohen J. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* 1960;20(1):37-46.

This page intentionally left blank

5. Quality, Safety, and Patient Outcomes

This page intentionally left blank

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-529

Design of a Visual Interface for Comparing Antibiotics Using Rainbow Boxes

Rosy TSOPRA[a,1], Shérazade KINOUANI[b,c], Alain VENOT[a], Marie-Christine JAULENT[a], Catherine DUCLOS[a] and Jean-Baptiste LAMY[a]
[a]*LIMICS, INSERM UMRS 1142, Université Paris 13, UPMC Université Paris 6, Paris, France*
[b]*Département de médecine générale, Univ. Bordeaux, France*
[c]*INSERM, team HEALTHY, UMR1219, Bordeaux, France*

Abstract. Non-optimal prescriptions of antibiotics have a negative impact on patients and population. Clinical practice guidelines are not always followed by doctors because the rationale of the recommendations is not always clear and can be difficult to understand. In this paper, we propose a new approach consisting in presenting the properties of antibiotics for allowing doctors to compare them and choose the most appropriate one. For that, we used and extended rainbow boxes, a new technique for overlapping set visualization. We tested our approach on 11 clinical situations related to urinary infections, and assessed the simplicity, the interest and utility with 11 doctors. 10 of them found that this approach was interesting and useful in clinical practice. Further studies are needed to confirm this preliminary work.

Keywords. Antibiotic prescription, Clinical Decision Support System, Visual analytics, Usability

1. Introduction

Non-optimal prescriptions of antibiotics have a negative impact on patients (risk of complications) and population (risk of bacteria resistance). To help general practitioners (GPs) to prescribe the right antibiotic, national health authorities elaborate Clinical Practice Guidelines (CPGs) which are evidence-based document written by a group of experts [1]. To facilitate their use, they are implemented in CDSSs (Clinical Decision Support Systems) which display, for each clinical situation, the antibiotics recommended in CPGs [2]. However, GPs are reluctant to use such systems because they don't always understand the rationale of the recommendation [3].

Rather than displaying the recommendation of CPGs (*e.g.* "prescribe this drug"), we propose a new approach consisting in presenting information for allowing the comparison of the various possible therapeutic options (*e.g.* "drug A has a low efficacy, drug B has a risk of serious adverse effects"), and letting GPs make their own choice. In a previous work [4,5], we identified the 6 properties of antibiotics that were necessary for choosing the antibiotic during prescription for urinary infections, and we ordered them according to their degree of importance. We hypothesize that this information could

[1] Corresponding author, rosy.tsopra@aphp.fr, LIMICS INSERM U1142, 15 rue de l'école de médecine, Esc D, 2ème étage, 75006 Paris

be used by GPs to choose one antibiotic among all those which are indicated for the clinical situation.

Few studies have focused on the presentation and the visual comparison of drug properties, and all of them used simple tables. Examples include a spreadsheet-like tool for reviewing and authoring drug properties, and tables for comparing the contraindications and the adverse effects of a new drug to a reference drug [6]. However, these tables are often very big and don't provide a global overview.

In a recent work, we designed a new technique for overlapping set visualization [7], *rainbow boxes* [8], and we applied it to the comparison of drug properties (contraindications and adverse effects). Rainbow boxes display the element to be compared (*e.g.* drugs) in columns, and the sets (*e.g.* drug properties) in labelled rectangular boxes that cover all the columns corresponding to the elements in the set. Larger boxes are placed at the bottom and two boxes can be side-by-side as long as they do not cover the same columns. A box can have holes, if the elements in the set are not displayed in consecutive columns. We designed a heuristic algorithm to find a near-optimal column order minimizing the number of holes, in a satisfying time.

The aim of the study was to design a visual interface displaying antibiotic properties and their importance, by using and adapting rainbow boxes. This interface could then be implemented in a CDSS for empiric antibiotic prescription [9]. In this paper, we will first describe the methods we followed for constituting a knowledge base containing the properties of antibiotics, for extending rainbow boxes and for evaluating the system. Then we will present the resulting interface and evaluation results and conclude.

2. Materials and Methods

2.1. Constitution of a Knowledge Base Containing the Properties of Each Antibiotic

By analyzing CPGs, we noticed that the reasoning for finding the appropriate antibiotic was based on the progressive exclusion of the antibiotics with the most important disadvantages. Therefore, we considered the 6 antibiotic properties as 6 potential disadvantages: (disadvantage #1) the antibiotic has a moderate clinical efficacy, (#2) the administration protocol is not convenient, (#3) it promotes the emergence of bacteria resistance, (#4) it is associated with a risk of serious or frequent adverse effects, (#5) it has a broad activity spectrum, (#6) it belongs to a precious class to keep for more serious indications. For each clinical situation, and for each antibiotic, a disadvantage is a Boolean value (true: the disadvantage is present, false: it is absent).

An algorithm developed in a previous work [4,5] was improved with the help of 6 medical experts to order the disadvantages of antibiotics according to their degree of importance. Disadvantage #1 (limited efficacy) was ranked as the most important, then #2, then #3, then #4-5-6 (same rank for the three). This order was transformed into a score computation, by associating a weight W_I with each disadvantage #I (from #1 to #6) (See details in results section). For a given clinical situation and antibiotic, the score is the sum of W_I for each disadvantage #I that is present. Thus, the antibiotic with the lowest score should be preferred.

From CPGs, we extracted all the clinical situations related to urinary infections. For each of them, we built a knowledge base containing the Boolean values for the 6 disadvantages of all the antibiotics that could be prescribed. The knowledge base was structured as an OWL formal ontology of the *ALIF* family of description logics.

2.2. Extending Rainbow Boxes to Display the Weight of Disadvantages

We extended the original rainbow boxes to take into account the weight associated with each disadvantage: we fixed the height of the box for disadvantage #I to a value proportional to its weight W_I. Consequently, the score can be visually computed by summing the height of all boxes present in a given column. We coloured the boxes in red. In addition, the saturation of the colour was also proportional to the weight (*i.e.* more important disadvantages have brighter red colours). We named the resulting visualization technique *weighted* rainbow boxes.

The OwlReady module for ontology-oriented programming [10] was used for accessing the knowledge base and generating rainbow boxes.

2.3. Evaluation Methods

Two sets of rainbow boxes were shown to 11 GPs: one with few antibiotics, and one with many antibiotics to compare. For each, we asked them 5 questions related to the simplicity, interest and utility of the interface (graduated with a Likert scale), and one question concerning the antibiotic they would prescribe. We also asked them if they would like to have such a system in clinical practice.

3. Results

3.1. Presentation of the Visual Interface for Comparing Antibiotics

11 clinical situations related to urinary infections were implemented in the form of rainbow boxes. The weights of the disadvantages were chosen to have a lexicographic order between disadvantages #1, #2, #3 and the conjunction of #4, #5 and #6; and an equal weight between #4, #5 and #6, *i.e.* $W_1 > W_2 + W_3 + W_4 + W_5 + W_6$; $W_2 > W_3 + W_4 + W_5 + W_6$; $W_3 > W_4 + W_5 + W_6$ and $W_4 = W_5 = W_6$. A lexicographic order means that disadvantage #1 is considered first, then #2 is considered in case of equality for #1, and so on. We arbitrarily chose the following values, which satisfy the previous formula: $W_{\#1} = 16.0$, $W_{\#2} = 7.9$, $W_{\#3} = 3.8$, and $W_{\#4} = W_{\#5} = W_{\#6} = 1.0$.

Figure 1 shows the resulting weighted rainbow boxes for two clinical situations, with 5 and 10 antibiotics. Antibiotics are displayed in columns, with their classes in headers, and the 6 disadvantages in the boxes. Only one hole is present (in the bottom, for ceftriaxone and the adverse effect disadvantages). The visualization allows to visually sum the weights of the various disadvantages for each antibiotic, *i.e.* to compute the score, easily, despite the presence of holes and gaps between some boxes. It is also very easy to find visually the antibiotics with the fewest and less important disadvantages (*e.g.* nitrofurantoin, for the second clinical situation in Figure 1). Moreover, they correspond to antibiotics recommended in rank 1 in CPGs.

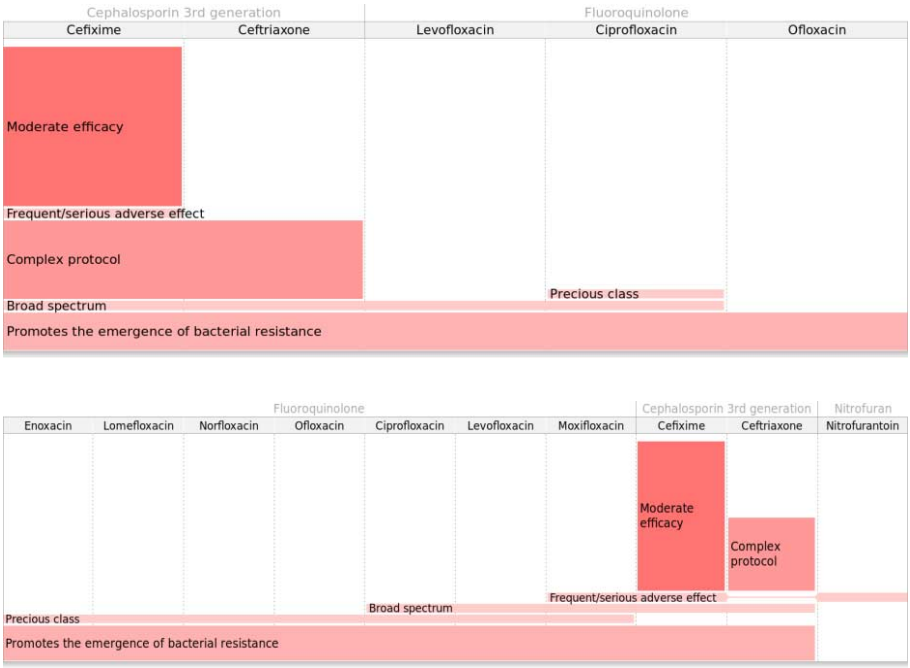


Figure 1. Weighted rainbow boxes showing the available antibiotics and their disadvantages in 2 clinical situations: pyelonephritis (top) and cystitis with risk of complication (bottom).

3.2. Evaluation Results

Rainbow boxes were perceived as easy for 64% of GPs in pyelonephritis, and 27.5% of GPs in cystitis. The reading of disadvantages looks easy for 73% of GPs in pyelonephritis and 64% of GPs in cystitis. About 90% of GPs found that the display of disadvantages of antibiotics could be interesting in clinical practice, and more than 91% of them found it could be useful in clinical practice. 5 GPs would like to have this system in clinical practice *vs* 3 who would not. In both clinical situations, 8 GPs prescribed the antibiotic with the fewest disadvantages displayed.

Table 1. Result of the evaluation. 11 GPs assessed two interfaces for cystitis and pyelonephritis

		Don't agree at all	Don't agree	Neutral	Agree	Fully agree
I find the interface easy	pyelonephritis	0	2	2	6	1
	cystitis	0	5	3	3	0
The disadvantages of each antibiotic are easy to read	pyelonephritis	0	2	1	6	2
	cystitis	0	2	2	6	1
I find interesting to visualize the disadvantages of antibiotics	pyelonephritis	0	0	1	4	6
	cystitis	0	0	1	5	5
It is useful to visualize the disadvantages of antibiotics in clinical practice	pyelonephritis	0	0	0	5	6
	cystitis	0	0	1	4	6
I would like to have this system in clinical practice		1	2	3	4	1

4. Discussion and Conclusion

Instead of displaying the recommendations of CPGs, we proposed a new approach consisting in comparing antibiotics according to their disadvantages, using rainbow boxes. Rainbow boxes were adapted to display the weight of each disadvantage. Although the rainbow boxes seemed to be more complex when the number of antibiotics is high, GPs still found the disadvantages easy to read. They also found that the comparison of antibiotics was interesting and useful.

Our approach allows GPs to choose the antibiotic to prescribe among all those able to treat the patient. It has several advantages. First, it allows GPs to adapt their prescription according to the patient profile [4] (*e.g.* if the most appropriate antibiotic cannot be given, then GP can choose a second one among those remaining). Second, the update of the system is facilitated, since the disadvantages could be updated from different resources (*e.g.* "protocol administration" from a drug database), instead of waiting many years for an updated version of the CPGs. Third, the use of weighted rainbow boxes allows GPs to see at a glance the antibiotic with the least disadvantages. Further studies are needed to confirm this preliminary work. Other types of infections (*e.g.* respiratory) should be tested. A broader evaluation with more GPs and with questions about the reasons why GPs would like to have a such system or not, needs to be conducted. The genericity of the approach to other domains beyond antibiotherapy, such as chronic disorders, needs also to be proven (other parameters (*e.g.* patient comorbidities) should probably be taken into account).

Acknowledgements

This work was funded by ANSM – Agence Nationale de Sécurité du Médicament et des produits de santé – AAP 2016 – RaMiPA

References

[1] Woolf SH, Grol R, Hutchinson A, Eccles M, Grimshaw J. Potential benefits, limitations, and harms of clinical guidelines. BMJ. 1999;318(7182):527-30.
[2] Sintchenko V, Coiera E, Gilbert GL. Decision support systems for antibiotic prescribing. Curr Opin Infect Dis. 2008;21(6):573-9.
[3] Varonen H, Kortteisto T, Kaila M. What may help or hinder the implementation of computerized decision support systems (CDSSs): a focus group study with physicians. Fam Pract. 2008;25(3):162-7.
[4] Tsopra R, Venot A, Duclos C. An Algorithm Using Twelve Properties of Antibiotics to Find the Recommended Antibiotics, as in CPGs. AMIA Annu Symp Proc. 2014;2014:1115-24.
[5] Tsopra R, Venot A, Duclos C. Towards evidence-based CDSSs implementing the medical reasoning contained in CPGs: application to antibiotic prescription. Stud Health Technol Inform. 2014;205:13-7.
[6] Wroe C, Solomon W, Rector A, Rogers J. DOPAMINE: a tool for visualizing clinical properties of generic drugs. Proc Fifth Workshop Intell Data Anal Med Pharmacol IDAMAP. 2000;61-65.
[7] Alsallakh B, Micallef L, Aigner W, Hauser H, Miksch S, Rodgers P. Visualizing Sets and Set-typed Data: State-of-the-Art and Future Challenges. Eurographics Conf Vis EuroVis. 2014;
[8] Lamy JB, Berthelot H, Favre M. Rainbow boxes : a technique for visualizing overlapping sets and an application to the comparison of drugs properties. Int Conf Inf Visuzalization. 253-260:2016.
[9] Tsopra R, Jais J-P, Venot A, Duclos C. Comparison of two kinds of interface, based on guided navigation or usability principles, for improving the adoption of computerized decision support systems: application to the prescription of antibiotics. J Am Med Inform Assoc JAMIA. 2014;21(e1):e107-116.
[10] Lamy J-B. Ontology-Oriented Programming for Biomedical Informatics. Stud Health Technol Inform. 2016;221:64-8.

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-534

A Method for Estimating the Risk Associated with Delaying Initial Treatment in Breast Cancer

Jonathan LENCHNER [a,1], Charity WAYUA [a]
[a] *IBM Research Africa, Catholic University of East Africa, Nairobi, Kenya*

Abstract. Over the past twenty-five years the time from diagnosis of breast cancer to the initiation of therapy has steadily grown. In this note we present a mechanism to give a ballpark estimate of the risk associated with delaying therapy given a specific set of presenting patient data.

Keywords. breast cancer, delay, initial treatment, risk

1. Introduction

When a patient is diagnosed with breast cancer, especially when the tumors are non-palpable or small, surgeons, oncologists and nurses will generally not stress the importance of prompt treatment. At the same time, the interval from diagnosis to initiation of therapy for breast cancer has been rising [1]. Should we be concerned?

Virtually all older studies about delaying therapy have been on the impact of delaying adjuvant chemotherapy [2,3,4,5]. An inconclusive 2013 study on the effect of surgical delays in Britain [6] is an exception. Very recent comprehensive studies, however, point to the fact that putting off therapy, both surgery and adjuvant chemotherapy, does in fact reduce one's survival probability [7,8]. Until now, however, it has not been possible to give an estimate of the risk associated with delaying therapy in the case of a particular set of presenting attributes (e.g. size of tumor, protein receptor status [ER, PR and HER2], single or multi-focality, age of patient, etc.). We present such a method, and provide our results in terms of the hazard ratio of 10-year mortality attributable to an incremental month of delaying therapy. Hazard ratio is defined as the ratio of the percentage of individuals incurring the hazard (in this case death in 10 years), given assumption of the risk (i.e. delay in therapy by one month), to the percentage of individuals incurring the hazard who have not assumed the risk.

2. Related Work

As far back as 1907, famed American surgeon William Stewart Halsted, responsible for, among other things, the introduction of the radical mastectomy for breast cancer treat-

[1] Corresponding Author; E-mail: jonathan.lenchner@ke.ibm.com

ment, appreciated the need for urgent treatment, stating [9] "Fortunately we no longer need the proof which our figures so unmistakably give that the slightest delay is dangerous and that, other things being equal, the prognosis is quite good in the early stage of breast cancer, two in three being cured, and bad, three in four succumbing, when the axillary glands are demonstrably involved."

Until very recently, however, systematic studies of the effect of postponing surgery or other primary treatments have not been performed, while times from diagnosis to treatment have increased [1]. Especially troubling are the results of [10] showing drastically different times to treatment based on racial/ethnic background and the patient's access to private insurance (points also reinforced in [8]). Further, given that the current authors are from sub-Saharan Africa, we call particular attention to the fact that there are noticeably longer times from diagnosis to treatment in the developing world compared to the developed world [11].

While studies of the effect of postponing surgery or other primary treatments have, until very recently, not been performed, there have been a number of modest size retrospective studies (population sizes ranging from 1788 to 6827) that have looked at the impact of delaying adjuvant chemotherapy following surgery [2,3,4,5]. While these studies have generally found a correlation between delay and poorer outcomes, one of these studies [5], that of 2782 Spanish patients from the time period 1990 to 1997, found no difference in either disease-free survival or 5-year overall survival based on time of initiation of adjuvant chemotherapy following surgery.

In 2013 a first study [6] of 53,689 female British breast cancer patients surprisingly found no statistically significant effect due to delaying surgery of up to 62 days on 5-year survival. In 2016, however, an editorial [12] and two very large prospectively collected studies [7,8] appeared in JAMA Oncology, giving a clear indication of the risk associated with either delaying surgery following initial diagnosis [7], or of delaying adjuvant chemotherapy following surgery [8]. The first, Bleicher et al. study [7], looked at data from 94,544 breast cancer patients from the Surveillance, Epidemiology and End Results (SEER)-Medicare Database, along with data from 115,790 breast cancer patients from the National Cancer Database. The authors analyzed 5-year overall survival and found an increased hazard ratio for mortality of 1.09 for each 30-day delay in treatment, up to 180 days of delay in the SEER-Medicare population and an increase hazard ratio for mortality of 1.10 for each 30-day delay in the National Cancer Database population. The second of the very recent studies, by Chavez-MacGregor et al. [8], looked at 24,843 breast cancer patients from the California Cancer Registry and found that delaying adjuvant chemotherapy after surgery by more than 90 days following surgery, compared to beginning in the first 30 days following surgery conferred a hazard ratio for 5-year mortality of 1.34 (and a hazard ratio for breast cancer-related 5-year mortality of 1.27).

Our risk assessment method uses a nomogram developed by researchers at the Memorial Sloan Kettering (MSK) Cancer Center in New York for assessing the likelihood of sentinel lymph node metastasis [13] given pathological characteristics of the primary tumor(s) and age of the patient. The nomogram was built based on a study population of 3786 breast cancer patients with sentinel lymph node biopsies performed at MSK between 1996 and 2002. The model was subsequently validated based on a second population of 1,545 patients receiving sentinel lymph node biopsies at MSK between 2002 and 2004 [14].

3. Our Risk Assessment Tool

The idea of our "back of the envelope" risk assessment tool is to use the MSK nomogram [13] to back into an estimate of the increased risk of delaying treatment by a given increment of time following initial diagnosis. In conjunction with the nomogram we use the comprehensive 2008 Soerjomataram study [15] that assesses the 10-year mortality-based hazard ratios associated with various prognostic factors.

Given a patient with a presenting case of breast cancer, the first step in assessing the patient's risk associated with delaying therapy is to fill in the patient's tumor and demographic information in the MSK nomogram. See the illustrative case on the left hand side of Figure 1 of a 56 year-old patient with multi-focal HER2+ breast cancer. The patient presented with a tumor of size 1.2 cm, another of size 1.0 cm, and a third, slightly smaller site of DCIS (ductal carcinoma in situ). After filling out the nomogram, we use

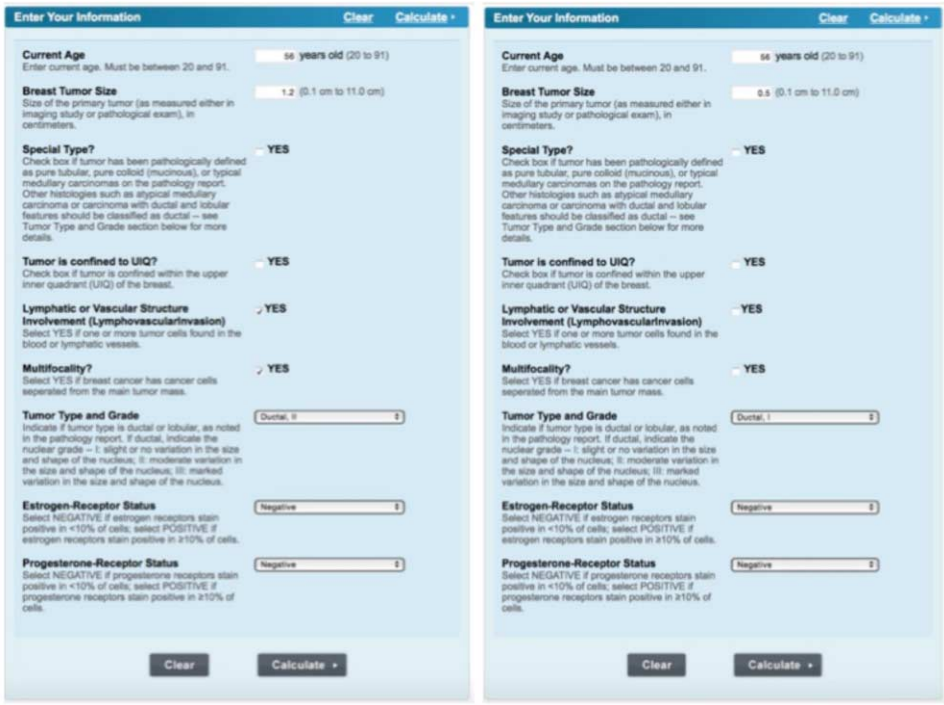


Figure 1. Left side: The completed nomogram for a patient with multi-focal HER2+ breast cancer. Right side: Backing up to the likely initial presentation of an earliest possible mammogram detection.

it to calculate the probability of sentinel node involvement, obtaining a probability that we denote by p and in this case $p = 0.56$. We then make note of the number of months since the patient has been in for their last mammogram and call this duration D. In this case, the patient's last mammogram was 14 months earlier so $D = 14$ months.

Next, we try to roll back the camera to the earliest time the exiting cancer might have been detectable under mammography. What would the original presentation have been like? In the case of our 56 year old patient we estimate that initial presentation to have been a single 0.5 cm instance of DCIS with no lymphatic or vascular structure involvement. The completed nomogram is given on the right hand side of Figure 1. We again calculate the probability of sentinel node involvement, obtaining a probability that we denote by p_0. In the case of our illustrative patient $p_0 = 0.07$. We now take the value of $D = 14$ months since the last mammogram and make the naive estimate that the first visible lesion would have manifested itself at a point $\frac{D}{2} = 7$ months earlier. Thus, over the estimated 7 months, the patient has incurred an additional likelihood of sentinel node involvement of $p - p_0 = 0.49$, and, extrapolating linearly, waiting an additional month will incur an incremental risk of approximately $\frac{p-p_0}{D/2} = \frac{0.49}{7} = 0.07$. Note that this linear addition of incremental risk cannot continue forever, since the risk of sentinel node involvement cannot exceed 1 and the delay vs. risk curve is almost certainly sigmoid. However, for values of p below about 0.75 the linearity assumption seems like a reasonably conservative one to make.

We next refer to [15] to get the hazard ratio for 10 year survival associated with sentinel lymph node involvement – the hazard ratio in this case being 2.4. We are incurring .07 of this hazard ratio in an incremental month, so that the expected hazard ratio of waiting an additional month is $.93(1) + .07(2.4) = 1 + 0.07 * (2.4 - 1.0) = 1.098$, or in this case, a value that is very similar to the aggregate value obtained in [7] when studying patients of all different tumor presentations and demographic characteristics. In general for values of p, p_0 and D, the approximate incremental hazard ratio of waiting an additional month, $\mathrm{HR}_{\text{inc-month}}$, is

$$\mathrm{HR}_{\text{inc-month}} = 1 + 2.8\frac{p - p_0}{D},$$

or for a generic reporting of hazard ratio associated with sentinel node involvement, $\mathrm{HR}_{\text{sent-node}}$,

$$\mathrm{HR}_{\text{inc-month}} = 1 + \frac{2(p - p_0)}{D}(\mathrm{HR}_{\text{sent-node}} - 1).$$

4. Conclusions and Next Steps

We have described a "back of the envelope" style estimate for approximating the risk of delaying treatment following initial diagnosis of breast cancer. A weakness of our method is that the risk of nodal involvement is supposed to increase linearly from a hypothetical first possible diagnosis, a supposition that is not a great one in case the probability of sentinel node involvement is already high. In this case it is better to make an explicit sigmoid assumption about the shape of the delay vs. risk curve and consider the conditional probability of additional, non sentinel node involvement, given the likely sentinel node involvement. Such additional node involvement becomes the predominant risk factor in such cases. MSK has a second breast cancer nomogram that can be used in this case [16]. In future work we will describe this more detailed calculation.

References

[1] R. J. Bleicher, K. Ruth, E. R. Sigurdson, E. Ross, Y.-N. Wong, S. A. Patel, M. Boraas, N. S. Topham, and B. L. Egleston, "Preoperative delays in the us medicare population with breast cancer," *Journal of Clinical Oncology*, vol. 30, no. 36, pp. 4485–4492, 2012.

[2] M. Colleoni, M. Bonetti, A. S. Coates, M. Castiglione-Gertsch, R. D. Gelber, K. Price, C.-M. Rudenstam, J. Lintner, J. Collins, B. Thrlimann, S. Holmberg, A. Veronesi, G. Marini, and A. G. for the International Breast Cancer Study Group, "Early start of adjuvant chemotherapy may improve treatment outcome for premenopausal breast cancer patients with tumors not expressing estrogen receptors," *JAMA Oncology*, vol. 18, no. 3, pp. 584–590, 2000.

[3] D. de M. Gagliato, A. M. Angulo-Gonzalez, X. Lei, R. L. Theriault, S. H. Giordano, V. Valero, G. N. Hortobagyi, and M. Chavez-MacGregor, "Clinical impact of delaying initiation of adjuvant chemotherapy in patients with breast cancer," *Journal of Clinical Oncology*, vol. 32, no. 8, pp. 735–744, 2014.

[4] C. Lohrisch, C. Paltiel, K. Gelmon, C. Speers, S. Taylor, J. Barnett, and I. A. Olivotto, "Impact on survival of time from definitive surgery to initiation of adjuvant chemotherapy for early-stage breast cancer," *Journal of Clinical Oncology*, vol. 24, no. 30, pp. 4888–4894, 2006.

[5] C. J. Sánchez, A. Ruiz, M. Martin, A. Antón, B. Munárriz, A. Plazaola, J. Schneider, P. M. del Prado, E. Alba, and A. Fernández-Aramuro, "Influence of timing of initiation of adjuvant chemotherapy over survival in breast cancer: a negative outcome study by the spanish breast cancer research group (geicam)," *Breast Cancer Research and Treatment*, vol. 101, no. 2, pp. 215–223, 2007.

[6] M. T. Redaniel, R. M. Martin, C. Cawthorn, J. Wade, and M. Jeffreys, "The association of waiting times from diagnosis to surgery with survival in women with localised breast cancer in England," *British Journal of Cancer*, vol. 109, pp. 42–49, 2013.

[7] R. J. Bleicher, K. Ruth, E. R. Sigurdson, J. R. Beck, E. Ross, Y.-N. Wong, S. A. Patel, M. Boraas, E. I. Chang, N. S. Topham, and B. L. Egleston, "Time to surgery and breast cancer survival in the united states," *JAMA Oncology*, vol. 2, no. 3, pp. 330–339, 2016.

[8] M. Chavez-MacGregor, C. A. Clarke, D. Y. Lichtensztajn, and S. H. Giordano, "Delayed initiation of adjuvant chemotherapy among patients with breast cancer," *JAMA Oncology*, vol. 2, no. 3, pp. 322–329, 2016.

[9] W. S. Halsted, "The results of radical operations for the cure of carcinoma of the breast," *Annals of Surgery*, vol. 46, no. 1, pp. 1–19, 1907.

[10] R. T. Bustami, D. B. Shulkin, N. ODonnell, and E. D. Whitman, "Variations in time to receiving first surgical treatment for beast cancer as a function of racial/ethnic background: a cohort study," *Journal of the Royal Society of Medicine*, vol. 5, no. 7, pp. 1–8, 2014.

[11] A. G. Q. Freitas and M. Weller, "Patient delays and system delays in breast cancer treatment in developed and developing countries," *Ciênc. saúde coletiva*, vol. 20, no. 10, pp. 3177–3189, 2000.

[12] G. Waks, T. A. King, and E. P. Winer, "Timeliness in breast cancer treatment the sooner, the better," *JAMA Oncology*, vol. 2, no. 3, pp. 302–304, 2016.

[13] M. S. K. C. Center, "Breast cancer nomogram: Sentinel Lymph Node Metastasis." `http://nomograms.mskcc.org/breast/BreastSLNodeMetastasisPage.aspx`. Accessed: 2016-10-30.

[14] J. L. B. Bevilacqua, M. W. Kattan, J. V. Frey, H. S. C. III, P. I. Borgen, and K. J. V. Zee, "Doctor, what are my chances of having a positive sentinel node? A validated nomogram for risk estimation.," *Journal of Clinical Oncology*, vol. 25, no. 24, pp. 3670–3679, 2007.

[15] I. Soerjomataram, M. W. J. Louwman, J. G. Ribot, J. A. Roukema, and J. W. W. Coebergh, "An overview of prognostic factors for long-term survivors of breast cancer," *Breast Cancer Research and Treatment*, vol. 107, no. 3, pp. 309–330, 2008.

[16] M. S. K. C. Center, "Breast cancer nomogram: Breast Additional Non SLN Metastases." `http://nomograms.mskcc.org/breast/BreastAdditionalNonSLNMetastasesPage.aspx`. Accessed: 2016-10-30.

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-539

A Standardized and Data Quality Assessed Maternal-Child Care Integrated Data Repository for Research and Monitoring of Best Practices: A Pilot Project in Spain

Carlos SÁEZ[a,b,1], David MONER[a,b], Ricardo GARCÍA-DE-LEÓN-CHOCANO[c], Verónica MUÑOZ-SOLER[b,c], Ricardo GARCÍA-DE-LEÓN-GONZÁLEZ[c], José Alberto MALDONADO[a,b], Diego BOSCÁ[a,b], Salvador TORTAJADA[a,e], Montserrat ROBLES[a], Juan M GARCÍA-GÓMEZ[a], Manuel ALCARAZ[c], Pablo SERRANO[d], José L BERNAL[d], Jesús RODRÍGUEZ[d], Gerardo BUSTOS[d], and Miguel ESPARZA[b]
[a] *Instituto Universitario de Tecnologías de la Información y Comunicaciones. Universitat Politècnica de València. Camino de Vera s/n. 46022 Valencia, España*
[b] *VeraTech for Health S.L., Valencia, España*
[c] *Hospital Virgen del Castillo, Yecla, España*
[d] *Hospital 12 de Octubre, Madrid, España*
[e] *Insituto de investigación sanitaria La Fe, Hospital Universitari i Politècnic La Fe, Valencia, España*

Abstract. We present the results of a pilot project of the Spanish Ministry of Health, Social Services and Equality, envisaged to the development of a national integrated data repository of maternal-child care information. Based on health information standards and data quality assessment procedures, the developed repository is aimed to a reliable data reuse for (1) population research and (2) the monitoring of healthcare best practices. Data standardization was provided by means of two main ISO 13606 archetypes (composed of 43 sub-archetypes), the first dedicated to the delivery and birth information and the second about the infant feeding information from delivery up to two years. Data quality was assessed by means of a dedicated procedure on seven dimensions including completeness, consistency, uniqueness, multi-source variability, temporal variability, correctness and predictive value. A set of 127 best practice indicators was defined according to international recommendations and mapped to the archetypes, allowing their calculus using XQuery programs. As a result, a standardized and data quality assessed integrated data respository was generated, including 7857 records from two Spanish hospitals: Hospital Virgen del Castillo, Yecla, and Hospital 12 de Octubre, Madrid. This pilot project establishes the basis for a reliable maternal-child care data reuse and standardized monitoring of best practices based on the developed information and data quality standards.

Keywords. Integrated data repositories, Data quality, Normalization, ISO 13606, Archetypes, Best practices, Quality indicators, Data reuse

[1] Corresponding autor: Carlos Sáez (carsaesi@ibime.upv.es), Instituto ITACA, Universitat Politècnica de València, Camino de Vera s/n, 46022, Valencia, Spain / VeraTech for Health SL, Avenida del Puerto 237-1, 46011, Valencia, Spain

1. Introduction

Integrated Data Repositories (IDRs) are becoming an essential resource enabling the biomedical data reuse on larger amounts and sources of data [1]. Several initiatives have been carried out on IDRs to provide access to biomedical research data, either as federated query tools [2,3] or as centralized repositories [4]. In most solutions, the adoption of a common data format was key. However, to our knowledge, the use of specific health information standards was limited [5]. Besides, it is agreed that the reliability of data reuse depends greatly on its Data Quality (DQ) [6]. Certainly, DQ assessment is considered a key component to any IDR [7], where some successful examples can be found in the recent literature [4,8].

We present the results of a pilot project of the Spanish Ministry of Health, Social Services and Equality (2015/07PN0010), envisaged to the development of a National IDR of maternal-child care information. The project had two main motivations. First, the evaluation of maternal and child health strategies [9,10] with the ultimate aim of disseminating best practices (BPs). Second, to provide a repository for population-based research, with a special focus on breastfeeding, as one of the main determinants for maternal and child health [10]. An IDR was developed as solution, which, based on health information standards, ensured a common interface for monitoring BPs of different hospitals and regions, and having its DQ assessed ensures a reliable data reuse.

2. Materials and Methods

The pilot was divided in three main workpackages: (1) definition of clinical information models and standardization of data, (2) DQ assessment and (3) definition of a proposal of BPs indicators. Clinical and data support was provided by the two participating hospitals: Virgen del Castillo Hospital, Yecla (VCH) and 12 de Octubre Hospital, Madrid (12OH). Figure 1 shows the architecture of the proposed solution.

Regarding to standardization, we used ISO 13606 archetypes [11] to provide the IDR with an information model about the data structure (how data is organized) and the data constrains to be fulfilled (which values are valid). To create the required archetypes, a multidisciplinary group of professionals was arranged. A proper information modelling is crucial to ensure that the relevant clinical information will be available for the particular data reuse purposes. Hence, the archetype creation methodology included the identification of the main clinical data structures, the selection and aggrupation of relevant data items for the required clinical domains, the search of reusable archetypes, the creation of new archetypes or adaptation of existing ones, and their validation by clinical experts. Finally, archetypes were mapped to the maternal-child care data extracted from the original data sources, which were transformed into the ISO 13606 archetype-compliant documents. Archetype edition and data transformation were performed using LinkEHR Studio [12]. The IDR was implemented in ExistDB, and queries were defined using native XQuery language.

The DQ assessment was carried out with a dedicated procedure based on seven dimensions [13]. Data completeness (non-missing data, weighting obligatory and optative elements), consistency (conformance to schema rules) and uniqueness (non-replicated identifiers) were calculated according to the archetype conformance requirements and based on our previous studies [14]. The multi-source and temporal variability of data (degree of data concordance among different sources and over time)

were assessed based on our probabilistic DQ methods [15-17]. Data correctness accounted for the number of possibly anomalous records (multivariate outliers). Finally, the predictive value dimension measured the baseline dataset potential to predict the breastfeeding continuity at one month as the AUC of a Naïve Bayes classifier with a 10-fold cross-validation estimation, as a measure of the data usefulness for this task.

Lastly, the BPs workpackage included two tasks. First, the formal definition of a set of BPs indicators based on an extensive review of literature. And second, the operative definition of the indicators and mapping of their variables to the archetype information for their automatic calculation.

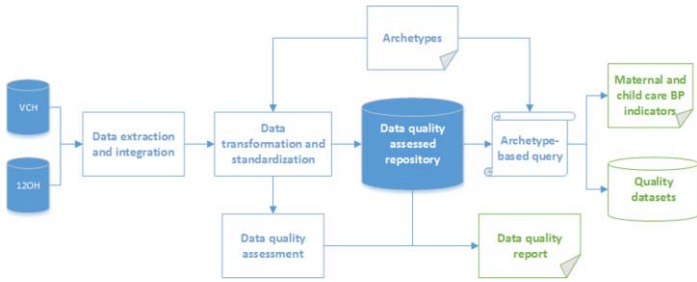


Figure 1. Architecture of the IDR solution for a standardized and reliable maternal-child care data reuse

3. Results

Two main archetypes were created as the basis for (1) a report of perinatal health information (Figure 2), including information about family history, gestation, delivery, birth, and maternity, and (2) a report of the infant feeding up to two years, including information about breastfeeding and the dates of introduction of different types of food. The two main archetypes were composed of 43 sub-archetypes. The two archetypes can be accessed at http://mm.linkehr.com/ (currently available in Spanish only).

To populate the IDR, the archetypes were mapped to the hospitals data sources and data were transformed to ISO 13606 instances. The VCH provided data from 3781 records for the newborn report and 2133 for the infant feeding report. The 12OH provided 1949 records for the newborn report, but only from their neonatal database.

DQ was assessed on the VCH and 12OH data separately for the perinatal and infant feeding datasets, evaluating the original EHR data (PRE) and after its standardization in the IDR (POST). A DQ report was generated, including the following main findings. Uniqueness: 100%, no replicated patient identifiers found. Completeness: in POST, completeness decreased due to stricter information requirements by the archetypes. Particularly, despite the higher completeness of the PRE 12OH neonatal dataset (77%), this filled a minor part of the more detailed perinatal archetype in POST (8%), whilst the average completeness of the VCH remained more stable (56% to 52%), filling in a higher degree the information required by the archetype. Consistency: in this pilot, due to the large amount of variables, data types and range checks were not included, thus, consistency results accounted for the un-conformance to data obligatoriness, and high measurements were obtained in general. Temporal stability: the method warned about a minor number of wrongly dated records in 12OH, showing up as an anomalous temporal subgroup; the VCH perinatal data showed three temporal subgroups (showing non-concordant data among their periods) related to two changes in the original EHR system; the VCH infant

feeding data was stable over time. Multi-source stability: a low stability between the VCH and 12OH perinatal data was found (0.08 out of 1), as expected due to their different populations (maternity vs neonatal units) and completeness derived from this reason. Correctness: an average of 1% of outlier records were found in all the datasets. Predictive value: an AUC of 0.60 was obtained using all the variables of VCH perinatal dataset. Six records did not pass the standardization process due to strong quality faults.

Finally, a set of 127 BPs indicators, grouped in six categories involving different clinical processes (from gestation to primary-care follow-up), was defined according to national and international recommendations (Euro-Peristat, WHO, UNICEF). All the variables in the operative definitions of indicators were mapped to the archetypes. A BP monitoring system was developed using XQuery programs on the standardized data.

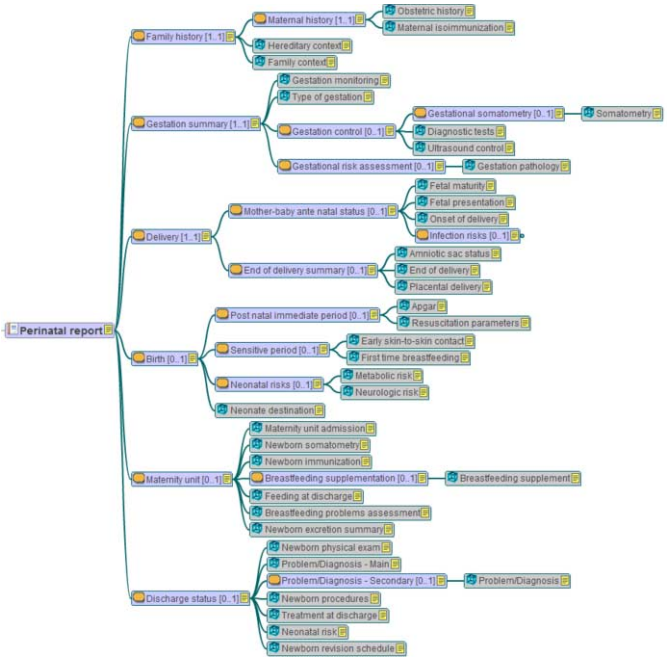


Figure 2. Contents of the perinatal report archetype

4. Discussion

The defined archetypes should be considered as an initial version, candidate to be revised by further professionals, towards a harmonized detailed clinical model. Further versions should also incorporate terminology bindings (e.g., to SNOMED-CT) to provide the semantics of the information structures and data values, not covered in this pilot. Mapping the local vocabularies of the two hospitals to the controlled vocabulary used in the archetypes was an intense task. Using standard terminologies in both the archetypes and the original data sources can solve this problem in the future.

The DQ assessment provided novel insights about the effect on DQ of data standardization using health information standards. Stricter information requirements can improve DQ in terms of completeness and consistency, and improve the usability of data given their contextualization. However, the change of the variable space given by the standardization must be considered in comparing PRE and POST measurements:

archetypes defined a larger set of information compared to the existing data. Consistency assessment can be improved including data type and range checks, what can be supported by the use of terminologies. The temporal variability results can be of utmost utility to support the mapping of data with variable representations over time.

The BPs monitoring based on standardized data would allow further enrolled hospitals getting instant BPs monitoring, comparison, and DQ assessment, just by providing equivalent standardized data. Finally, other repository technologies are to be explored to support advanced query needs and to improve the efficiency of the IDR.

5. Conclusion

This pilot project established the basis for a national IDR for a reliable maternal-child data reuse and standardized monitoring of BPs. The discussed lessons learned can facilitate the scaling-up of the project in national and international actions. The developed approach can be replicated in additional healthcare domains.

Acknowledgements

Work co-supported by grants: RTC-2014-1530-1, AEST/2016/023 and DI-14-06564.

References

[1] L. Toubiana and M. Cuggia, Big Data and Smart Health Strategies: Findings from the Health Information Systems Perspective. *IMIA Yearbook.* 2014;9(1):125–7.
[2] A.J. McMurry et al. SHRINE: Enabling Nationally Scalable Multi-Site Disease Studies. Carter KW, editor. *PLoS ONE*. 2013 7;8(3):e55811.
[3] G.M. Weber et al. Direct2Experts: a pilot national network to demonstrate interoperability among research-networking platforms. *J Am Med Inform Assoc*. 2011;18:157–60.
[4] S.N. Murphy et al. Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2). *J Am Med Inform Assoc*. 2010;17(2):124–30.
[5] K.L. Walker et al. Using the CER Hub to ensure data quality in a multi-institution smoking cessation study. *J Am Med Inform Assoc*. 2014;21(6):1129–35.
[6] N.G. Weiskopf and C. Weng, Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research, *J Am Med Inform Assoc,* 2013;20(1):144–51.
[7] S.L. MacKenzie et al, Practices and perspectives on building integrated data repositories: results from a 2010 CTSA survey, *J Am Med Inform Assoc* 2012;19(1):119–24.
[8] M.G. Kahn et al, A Pragmatic Framework for Single-site and Multisite Data Quality Assessment in Electronic Health Record-based Clinical Research: *Med Care*. 2012;50:S21–9.
[9] Estrategia de atención al parto normal y nacimiento en el Sistema Nacional de Salud. Observatorio de Salud de la Mujer y del Sistema Nacional de Salud. Ministerio de Sanidad y Consumo, 2008.
[10] Evidence for the Ten Steps to Successful Breastfeeding. Geneva: WHO; 1998.
[11] ISO 13606:2008 - Health informatics - Electronic health record communication. 2008.
[12] J.A. Maldonado et al, LinkEHR-Ed: a multi-reference model archetype editor based on formal semantics, *Int J Med Inf*, 2009:78;8:559–70.
[13] C. Sáez et al. Organizing data quality assessment of shifting biomedical data. *Stud Health Technol Inform*, 180:721–725, 2012.
[14] R. García-de-León-Chocano et al. Construction of quality-assured infant feeding process of care data repositories: Construction of the perinatal repository (Part 2). *Comput Biol Med*, 2016:71(1):214–22.
[15] C. Sáez et al. Applying probabilistic temporal and multi-site data quality control methods to a public health mortality registry in Spain: A systematic approach to quality control of repositories. *J Am Med Inform Assoc,* 2016;23:1085-95.
[16] C. Sáez et al. Stability metrics for multi-source biomedical data based on simplicial projections from probability distribution distances. *Stat Methods Med Res*. 2017:26(1):312–336.
[17] C. Sáez et al. Probabilistic change detection and visualization methods for the assessment of temporal stability in biomedical data quality. *Data Min Knowl Discov*. 2015;29(4):950–75.

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-544

Using the MRC Framework for Complex Interventions to Develop Clinical Decision Support: A Case Study

Dawn DOWDING[a,b1], Valentina LICHTNER[c] and S. José CLOSS[c]
[a] *Columbia University School of Nursing, USA*
[b] *Visiting Nurse Service of New York, USA*
[c] *University of Leeds, UK*

Abstract. The Medical Research Council (MRC) framework for complex interventions provides useful guidance to assist with the development and evaluation of health technology interventions such as decision support. In this paper we briefly summarise a project that focused on designing a decision support intervention to assist with the recognition, assessment and management of pain in patients with dementia in an acute hospital setting. We reflect on our experience of using the MRC framework to guide our study design, and highlight the importance of considering decision support interventions as complex interventions.

Keywords. Clinical decision support systems, pain measurement, pain management, dementia

1. Introduction

The Medical Research Council (MRC) guidelines for complex interventions [1] provide guidance to researchers on the process for developing and evaluating interventions that contain several interacting components. The aim of the framework is to ensure that interventions are empirically and theoretically founded, and that considerations are given both to the effectiveness of the intervention and *how* it works. In this paper we report on our experiences of using the MRC framework as the basis for developing a clinical decision support intervention, focusing on the assessment and management of pain in patients with dementia in an acute care setting.

1.1. Complex Interventions

Complex interventions have 'several dimensions of complexity' such as variations in the number of intervention components, behaviours and degree of flexibility required to implement it, the groups it targets and the interactions between components [1]. The MRC framework (figure 1) provides guidance on how to design and evaluate such interventions in a structured way. It highlights the importance of the development phase of intervention design, ensuring that there is an evidence base and theory to support the

[1] Corresponding author: Dawn Dowding PhD, RN, FAAN, Columbia University School of Nursing, 617 W 168th Street, New York, NY 10032, USA. E-mail: dd2724@columbia.edu

intervention, modelling both the intervention process and outcomes, before it is piloted for feasibility [1].

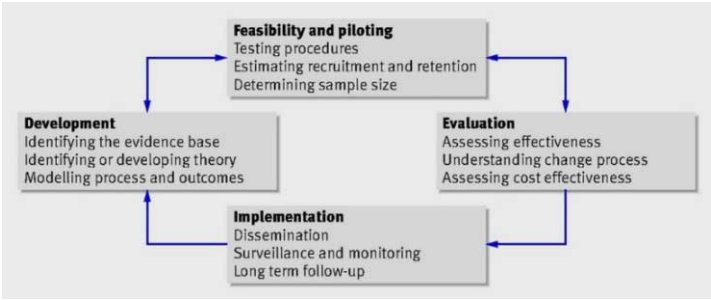


Figure 1. MRC framework of complex interventions

1.2. Pain in Patients with Dementia and Decision Support Interventions

Pain is a common symptom in older adults, and it is estimated that approximately 50% of people who have dementia also experience pain [8]. There is growing evidence that pain is often inadequately treated in patients with dementia; patients often have difficulty with recall, interpretation, identification and responses to pain making it challenging for health care professionals to evaluate their pain experiences [4]. These difficulties are often compounded in an acute care setting, where the environment may increase a person with dementia's sense of confusion and disorientation, and where staff may be unfamiliar with their individual pain responses [7].

Clinical decision support systems 'provide clinicians with patient-specific assessments or recommendations to aid clinical decision making' [3]. They often integrate information from a variety of sources using sophisticated technology and are implemented in a complex environment (that of a health care organisation with different layers of individual and social units collaborating together).

In this study we aimed to develop a decision support system that could assist clinicians with the complex task of identifying and treating pain in patients with dementia in an acute care setting. It focused on the first and second stages of the MRC framework; developing the intervention and exploring its feasibility.

2. Theory Development and Identifying the Evidence Base

Existing models of pain recognition, assessment and management of pain assume a linear process that could be compared to a linear judgement and decision making process (figure 2) [2], mirroring individual cognition.

This conceptual model was used as the starting point for our research, with the focus on identifying existing tools (that could potentially be used as the basis for our decision support intervention) and modelling the processes by which pain was recognised, assessed and managed in patients with dementia in acute care settings.

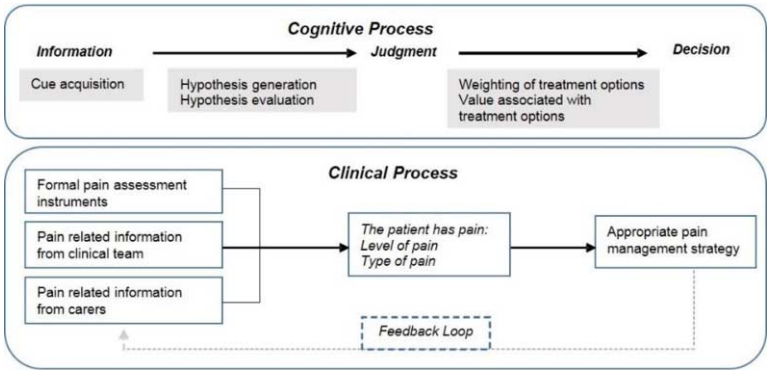


Figure 2: The cognitive and clinical process for recognising, assessing and managing pain [2]

3. Methods

A systematic review of systematic reviews of pain assessment tools, and a multiple case site study with embedded units of analysis. The full methods and results of these studies have been reported elsewhere [2; 5; 6]. Here we provide a brief overview and key results, to provide context for our discussion of the utility of the MRC framework to guide the intervention development.

3.1. A Systematic Review of Systematic Reviews of Pain Assessment Tools for Patients with Dementia [6]

Systematic reviews of pain assessment tools were identified through searching databases (e.g. Medline, Embase, Cochrane) the JBI Database of Systematic Reviews and the DARE database, alongside reference chaining. Reviews were included in our 'meta-review' if they included pain assessment tools involving adults with dementia/cognitive impairment and provided psychometric data on the tools evaluated. Each review was assessed for risk of bias using the AMSTAR critical appraisal tool. Data were extracted and summarised. At least two reviewers carried out each element of the review procedure (i.e. review inclusion/exclusion, assessment of methodologic quality, data extraction).

3.2. A Multiple Case Site Study Using Ethnography [2; 5]

Data were collected in four NHS hospital trusts across England and Scotland, varying in size and types of service provision. Wards within each hospital were theoretically sampled to provide an overview of care provided to patients with dementia in a variety of settings (e.g. orthopedic, acute medicine, elderly care). Data were collected using non-participant observation of care interactions, semi-structured interviews with clinical staff and informal carers and audits of patient notes for documentation related to the recognition, assessment and management of pain.

4. Results

Ten papers reporting the results of 8 reviews were included in the meta-review. Each review summarised between 8 and 13 pain assessment tools, providing data for 28 tools in total. Overall, there were limited data on the reliability, validity and clinical utility of any of the tools evaluated in the reviews. On the basis of the evidence, we were unable to identify one specific tool that could be used as the basis for our decision support intervention.

In our case study we observed 31 patients with dementia for a total of 170 hours of observation; we interviewed 52 health care staff and 4 carers. Our analysis highlighted the difficulties patients with dementia had communicating with staff about their pain. Patients with dementia had significant issues communicating pain verbally, and their interactions with staff were often brief, and rarely with the same person. These problems with communication affected clinicians' abilities to reassess pain following administration of therapy, and often affected whether a patient received medication at all. Overall the process of pain recognition, assessment and management involved 'putting a picture together' of a patient's pain, which required clinicians to share information across individuals, and through written documentation which was often fragmented and kept in professional 'silos'.

5. Discussion

The results of our study highlight the importance of considering theory and aspects of *how* an intervention may be thought to work in practice, as the first stage of intervention development. Our study used existing conceptualisations of how pain is thought to be identified, assessed and managed in clinical settings, underpinned by decision making theories. However the study results highlighted the need to refine that theory; if we had assumed that the focus of the decision tool should support a linear process focused on one clinician, we would not be reflecting the actual decision process we discovered through our ethnographic work. We have subsequently expanded the theoretical basis of our intervention to include an acknowledgement of the work of 'sense-making' in pain recognition, assessment and management [2] which can then provide the basis for an intervention that may actually have more utility for clinical staff in a practice setting.

What our work has highlighted is that the MRC framework provides a good starting point for intervention development, but as often the interventions we are developing are being implemented in complex environments, that process has to be cyclical and flexible to adapt to the environment and project findings. Our study also highlighted the issues related to summarising and evaluating existing evidence; the meta-review for our study was extremely complex and did not identify one 'best' tool that could be used in practice. On reflection this may be a frequent issue with systematic reviews of complex interventions, that the results tend to be also complex and context dependent. The MRC framework reminds us that interventions need to be evidence based, but finding conclusive evidence may be the first challenge.

Since the MRC framework was originally published and then revised (in 2006) the science of complex intervention development and evaluation has progressed considerably. Whilst we found the framework a useful starting point to provide a structure for our research study, the complexity of the theoretical and clinical environment suggest that other conceptualisations of intervention development and

evaluation may also be useful. For example realist evaluation methods, originally developed to explain how program interventions may work in one environment, but not another, may provide one way of supporting complex intervention development. This approach, where the researcher outlines and refines theories based on how individuals interact with the resources provided by an intervention (known as mechanisms) could enable a more flexible and reflexive approach both to theory development and testing in complex intervention research.

In general health informatics solutions, such as decision support interventions and other technological innovations, are rarely conceptualised in terms of their complexity. Given the complex nature of such technology, that individual users may interact with that technology in a number of ways, and the complexity of the environments where they are often introduced, we also believe that both the MRC framework and other approaches to evaluation could provide a useful framework for informatics researchers in the future.

6. Acknowledgements

The authors would like to thank the following members of the research team: Nick Allcock, Michelle Briggs, Anne Corbett, Philip Esterhuizen, Nita Gorasia, John Holmes, Claire Hulme, Kirstin James, John Keady, Andrew F Long, Liz McGinnis, John O'Dwyer, Liz Sampson, Caroline Swarbrick. This project was funded by the National Institute for Health Research HS&DR Programme (11/2000/05). The views and opinions expressed therein are those of the authors and do not necessarily reflect those of the HS&DR, NIHR, NHS or the Department of Health.

References

[1] Craig P, Dieppe P, Mcintyre S, Michie S, Nazareth I, and Petticrew M, Developing and evaluating complex interventions: new guidance, in, Medical Research Council, London, UK, 2006.

[2] Dowding D, Lichtner V, Allcock N, Briggs M, James K, Keady J, Lasrado R, Sampson EL, Swarbrick C, and Closs SJ, Using sense-making theory to aid understanding of the recognition, assessment and management of pain in patients with dementia in acute hospital settings, *International Journal of Nursing Studies* **53** (2016), 152-162.

[3] Kawamoto K, Houlihan CA, Balas EA, and Lobach DF, Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success, *British Medical Journal* **330** (2005), 765.

[4] Leong IY and Nuo TH, Prevalence of pain in nursing home residents with different cognitive and communicative abilities, *The Clinical Journal of Pain* **23** (2007), 119-127.

[5] Lichtner V, Dowding D, Allcock N, Keady J, Sampson EL, Briggs M, Corbett A, James K, Lasrado R, Swarbrick C, and Closs SJ, The assessment and management of pain in patients with dementia in hospital settings: a multi-case exploratory study from a decision making perspective, *BMC Health Services Research* **16** (2016), 427.

[6] Lichtner V, Dowding D, Esterhuizen P, Closs SJ, Long AF, Corbett A, and Briggs M, Pain assessment for people with dementia: a systematic review of systematic reviews of pain assessment tools, *BMC Geriatrics* **14** (2014), 138.

[7] Sampson EL, White N, Lord K, Leurent B, Vickerstaff V, Scott S, and Jones L, Pain, agitation, and behavioural problems in people with dementia admitted to general hospital wards: a longitudinal cohort study, *Pain* **156** (2015), 675-683.

[8] van Kooten J, Delwel S, Binnekade TT, Smalbrugge M, van der Wouden JC, Perez RSGM, Rhebergen D, Zuurmond WWA, Stek ML, Lobbezoo F, Hertogh CMPM, and S. EJA, Pain in dementia: prevalence and associated factors: protocol of a multidisciplinary study, *BMC Geriatrics* **15** (2015), 29.

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-549

Square² - A Web Application for Data Monitoring in Epidemiological and Clinical Studies

Carsten Oliver SCHMIDT[1], Christine KRABBE, Janka SCHÖSSOW, Martin ALBERS, Dörte RADKE, Jörg HENKE
Institute for Community Medicine SHIP-KEF, University Medicine of Greifswald, Walther Rathenau Str. 48, 17475 Greifswald

Abstract. Valid scientific inferences from epidemiological and clinical studies require high data quality. Data generating departments therefore aim to detect data irregularities as early as possible in order to guide quality management processes. In addition, after the completion of data collections the obtained data quality must be evaluated. This can be challenging in complex studies due to a wide scope of examinations, numerous study variables, multiple examiners, devices, and examination centers. This paper describes a Java EE web application used to monitor and evaluate data quality in institutions with complex and multiple studies, named Square². It uses the Java libraries Apache MyFaces 2, extended by BootsFaces for layout and style. RServe and REngine manage calls to R server processes. All study data and metadata are stored in PostgreSQL. R is the statistics backend and LaTeX is used for the generation of print ready PDF reports. A GUI manages the entire workflow. Square² covers all steps in the data monitoring workflow, including the setup of studies and their structure, the handling of metadata for data monitoring purposes, selection of variables, upload of data, statistical analyses, and the generation as well as inspection of quality reports. To take into account data protection issues, Square² comprises an extensive user rights and roles concept.

Keywords. data quality; data monitoring; web-application; epidemiology; statistical analyses

1. Introduction

To enable valid scientific inferences from epidemiological and clinical studies it is essential to obtain a high data quality. A broad scope of recommendations for study design, and quality assurance measures have been described to achieve this goal.[6; 10]. An indispensable aspect of quality assurance processes is a functioning data quality monitoring. Several data quality indicators have been described [1-3; 7; 8] for this purpose. They target data properties such as missing values, implausible values, extreme values, and measures of reliability and validity. For example, the "Guideline for the Adaptive Management of Data Quality in Cohort Studies and Registers" describes 51 quality indicators which are organized in the categories of plausibility (27 indicators), organization (16 indicators), correctness (6 indicators), and metadata (2 indicators).[5]

Data generating departments aim to detect data irregularities as early as possible to guide quality management processes. After the completion of data collections the quality

[1] Corresponding author, University Medicine Greifswald, Institute for Community Medicine, SHIP-KEF, Walther-Rathenau Str. 48. 17475 Greifswald; E-mail: Carsten.schmidt@uni-greifswald.de.

of this data must be rigorously evaluated before the start of scientific analyses.[4] In complex studies this can be a challenging task due to a wide scope of examinations with numerous study variables, as well as multiple examiners, devices, and examination centers. The Study of Health in Pomerania (SHIP) may serve as an example for this complexity.[9] SHIP studies the prevalence and incidence of risk factors, subclinical disorders, clinical diseases, and their associations. To date, almost 9000 adults have been examined up to five times. Examinations consisted of an extensive computer assisted personal interview, self-report questionnaires, the collection of biomaterials (blood, urine, faeces, saliva), imaging (e.g. ultrasound of the carotid artery, liver, thyroid, heart; full-body magnetic resonance imaging (MRI)), a dental and dermatological examination, and more. Thousands of variables, grouped into dozens of examination categories must be managed. The complexity is increased by changing examination teams, temporal examination centers and the conduct of numerous secondary data generation projects, which are for example related to the reading of MRI images (e.g. disc herniation). All of these secondary data collections may themselves be regarded as studies with an independent workflow. Other major cohort studies face comparable complexities.[11]

Under these circumstances, an efficient, manual data monitoring method seemed no longer feasible. Therefore, we aimed to standardize processes by developing appropriate IT tools. First, a partial standardization of the workflow was achieved by combining a manually controlled STATA analysis environment with a web frontend to generate PDF reports. This was implemented in the year 2010. Second, based on our experience with the first data quality analysis tool and our interactions with internal and external SHIP project partners, and team members we developed a web application to control the entire data monitoring process, Square², which is described in detail in this paper. The decision for a new development was made because (1) software solutions to monitor data quality were highly uncommon in major epidemiologic studies [3] despite their use in other fields of research, (2) existing solutions did not meet to a sufficient degree the requirements of large epidemiologic cohort studies. Important requirements were, among others, a standalone web-application allowing for a multi-study management with a differentiated rights- and roles concept to safeguard data protection issues, the possibility to automatically generate standard reports without statistical programming, the option to flexibly adapt reports to individual demands, a strong focus on measurement error related issues, flexible extension of statistical functionalities based on the integration of standard statistical packages (e.g. R), and non-commercial availability of all components to avoid additional costs for academic users.

2. Methods

Square² was designed as a Java EE web application, deployed in Tomcat 8. All data is stored in a PostgreSQL database. The statistics backend is R because of its free availability, wide acceptance in the statistical community, fast growing scope of packages, and the option to run it as a server process. LaTeX is used for the generation of print ready PDF reports. An overview of the components is provided in Figure 1. A GUI manages the entire workflow.

Square² uses the Java libraries Apache MyFaces 2, extended by BootsFaces for layout and style. RServe and REngine manage calls to R server processes. Additionally, we use unit testing libraries (JUnit). The following in-house developed libraries are used:

ShipDBM, a data persistence library, providing access of the web application to PostgreSQL, and Pwencrypt, a library for password encryption.

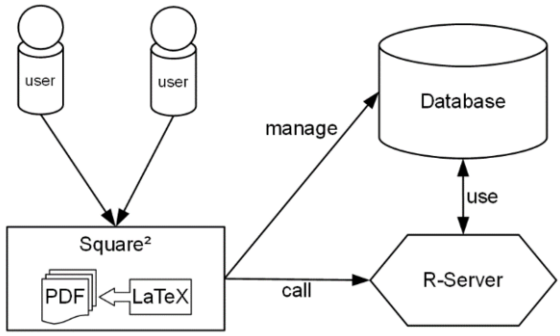


Figure 1. Square² Components.

R performs statistical analyses, and results are stored in the database. Analyses can be processed asynchronously to improve computational speed. The in-house developed R package squareControl manages all R server processes. It uses the package „rpg“ for persistence, "futile.logger" to log process information, and "parallel" to parallelize tasks.

Study data as well as all metadata and analyses results are stored in PostgreSQL. Square² starts calculations on the R server and disconnects thereafter, leaving all subsequent processes to the R Server. Thus all computational load remains on the R Server, including data base operations. All analysis results are stored in PostgreSQL. Graphical elements are embedded in LaTeX documents as base64 encodings.

3. Results

Square² is currently used within SHIP [9] but has been designed to meet the needs and requirements of different studies. The workflow consists of the following steps:

1. Study management: First, a new study needs to be defined in Square² for data monitoring purposes by providing descriptive information such as study name, study description, and if possible begin and end dates.
2. Study structure: In this step, the study structure and all necessary metadata for data monitoring are defined. Hierarchical elements of a study may consist of groups of examinations (e.g. medical examinations), examinations (e.g. hand grip, anthropometric measurements, blood pressure), and variables (e.g. the first measurement of hand grip strength in the left hand in kg). Next, metadata associated with a study and single variables are added. Metadata fields include, for example, the variable type (e.g. categorical, continuous, count), plausibility limits for continuous or count variables, reference categories for categorical variables, missing value indicators, observer, device or center indicators, and measurement times and dates. Elements of the study structure may either be added manually by using the GUI or by an import of available metadata.
3. Variable sets. Sets of variables are subsequently defined for data monitoring purposes. These sets may consist of variables from different studies. Quality

officers may only create reports for variables from variable sets to which they have been assigned.

4. Data management. This functionality controls the upload of study data for statistical analyses.
5. Statistics. The statistics module serves to enter R scripts into Square². For each R script, input and output parameters must be defined to properly link R scripts with variable and study metadata and to enable the web-application to integrate statistical output into reports. R-scripts are assigned to predefined report categories (e.g. descriptive statistics, missing values, extreme values, observer or device variability). Statisticians may add statistical functions without in depth knowledge of the web application.
6. Templates. The reporting of data quality within studies often follows standard requirements and reports should be highly comparable. For this purpose, templates can be created to structure reports (using elements such as headers, sub-headers, text blocks, tables, statistical output, page breaks).
7. Quality reports. The preceding steps provide the necessary background information to now create specific quality reports. First, a new report is defined by assigning a name, templates, and a variable set. Second, an analysis matrix is created to link variables from the variable set with statistical routines. Third, additional information may be entered into the predefined text fields. Fourth, the report is generated, and may subsequently be inspected.

Once all background information is entered, users mainly work in step seven. Square² includes an extensive user rights and roles concept. This includes tailored access to study data and reports by assigning personnel to specific studies, or even to specific sets of variables within studies to protect data safety. Multiple roles can be managed such as principal investigators (e.g. for the definition or deletion of studies), quality officers (e.g. to add and modify study metadata and to create reports), statisticians (e.g. to add R scripts), and examiners (e.g. to only read reports related to their own examinations).

4. Conclusion

Square² was primarily designed to support the monitoring of data quality in institutions running multiple complex studies. It allows for an efficient and timely generation of standard data quality reports. Efficiency is essential because funding for extensive data quality control activities is often limited. The design of Square² draws strongly from the concept of automatization of monitoring processes. However, there are limits. Square² may not be the most appropriate tool to address highly specific data quality aspects. This limitation is mainly related to logistic considerations. While the modular design of Square² allows for a strong degree of individualization, the integration of such highly specific analyses might be realized more quickly outside the Square² framework. The strengths of Square² are most apparent when there is a need for a repetitive reporting with a centralized demand to oversee and control these reporting activities.

Another limitation is related to the focus on the properties of measured data. However, the obtained data quality in a study may only be fully appreciated by interpreting results in light of additional study meta-information, such as the study design.

The development of Square² is still ongoing. Access to report contents may for example be based on HTML pages rather than PDF reports. We intend to classify R-

routines based on data quality indicators as described by the "Guideline for the Adaptive Management of Data Quality in Cohort Studies and Registers".[5]

Square² may be accessed for academic use through scientific cooperation projects, please contact the first author for this purpose. Free access for academic users to a web-based instance of Square² is projected to be available as part of an ongoing network project on data quality indicators in cooperation with the TMF (http://www.tmf-ev.de), an umbrella organization for networked medical research in Germany.

5. Acknowledgements

This work was supported by the Ministry for Education, Science and Culture of the State of Mecklenburg-Vorpommern, the European Social Fund (Grant UG 11 035A), and by the German Research Foundation (DFG, SCHM 2744/3-1).

References

[1] D.G. Arts, N.F. De Keizer, and G.J. Scheffer, Defining and improving data quality in medical registries: a literature review, case study, and generic framework, *Journal of the American Medical Informatics Association* **9** (2002), 600-611.

[2] J.R. Brestoff and J. Van den Broeck, Reporting Data Quality, in: *Epidemiology Principles and Practical Guidelines,* J. Van den Broeck and J.R. Brestoff, eds., Springer, Dordrecht, 2013.

[3] O. Harel, E.F. Schisterman, A. Vexler, and M.D. Ruopp, Monitoring quality control: can we get better data?, *Epidemiology* **19** (2008), 621-627.

[4] M. Huebner, W. Vach, and S. le Cessie, A systematic approach to initial data analysis is good research practice, *Journal of Thoracic and Cardiovascular Surgery* **151** (2016), 25-27.

[5] M. Nonnemacher, D. Nasseh, and J. Stausberg, *Datenqualität in der medizinischen Forschung: Leitlinie zum Adaptiven Datenmanagement in Kohortenstudien und Registern*, TMF e.V., Berlin, 2014.

[6] C.O. Schmidt, Anwendungsempfehlungen für Kohorten in: *Datenqualität in der medizinischen Forschung: Leitlinie zum Adaptiven Datenmanagement in Kohortenstudien und Registern.,* M. Nonnemacher, N. D., and S. J., eds., TMF e.V., Berlin, 2014, pp. 117-127.

[7] J. Van den Broeck, S.A. Cunningham, R. Eeckels, and K. Herbst, Data cleaning: detecting, diagnosing, and editing data abnormalities, *PLoS Medicine* **2** (2005), e267.

[8] D. Venet, E. Doffagne, T. Burzykowski, F. Beckers, Y. Tellier, E. Genevois-Marlin, U. Becker, V. Bee, V. Wilson, C. Legrand, and M. Buyse, A statistical approach to central monitoring of data quality in clinical trials, *Clinical Trials* **9** (2012), 705-713.

[9] H. Volzke, D. Alte, C.O. Schmidt, D. Radke, R. Lorbeer, N. Friedrich, N. Aumann, K. Lau, M. Piontek, G. Born, C. Havemann, T. Ittermann, S. Schipf, R. Haring, S.E. Baumeister, H. Wallaschofski, M. Nauck, S. Frick, A. Arnold, M. Junger, J. Mayerle, M. Kraft, M.M. Lerch, M. Dorr, T. Reffelmann, K. Empen, S.B. Felix, A. Obst, B. Koch, S. Glaser, R. Ewert, I. Fietze, T. Penzel, M. Doren, W. Rathmann, J. Haerting, M. Hannemann, J. Ropcke, U. Schminke, C. Jurgens, F. Tost, R. Rettig, J.A. Kors, S. Ungerer, K. Hegenscheid, J.P. Kuhn, J. Kuhn, N. Hosten, R. Puls, J. Henke, O. Gloger, A. Teumer, G. Homuth, U. Volker, C. Schwahn, B. Holtfreter, I. Polzer, T. Kohlmann, H.J. Grabe, D. Rosskopf, H.K. Kroemer, T. Kocher, R. Biffar, U. John, and W. Hoffmann, Cohort profile: the study of health in Pomerania, *Int J Epidemiol* **40** (2011), 294-307.

[10] C.W. Whitney, B.K. Lind, and P.W. Wahl, Quality assurance and quality control in longitudinal studies, *Epidemiologic Reviews* **20** (1998), 71-80.

[11] H.E. Wichmann, R. Kaaks, W. Hoffmann, K.H. Jöckel, K.H. Greiser, and J. Linseisen, [The German National Cohort], *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz* **55** (2012), 781-787.

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-554

Exploring the Notion of Hazards for Health IT

Ibrahim HABLI [a,1], Sean WHITE [b], Stuart HARRISON [b] and Manpreet PUJARA [b]
[a] *University of York*
[b] *NHS Digital*

Abstract. Safety analysis is centred on identifying a set of hazards that form the basis of risk assessment. In healthcare, hazards are potential sources of harm to patients and as such the risk of these has to be assessed and managed. With the increased reliance on Health IT systems in health and social care settings, some of these hazards are associated with the development and use of these systems. In this paper we examine current practices in hazard identification, focusing on how clinicians and engineers approach this task within the Health IT safety assurance process. We highlight certain technical and organisational challenges and discuss approaches to improving current practices and promoting learning initiatives.

Keywords. Digital health, health IT, hazards, patient safety

1. Introduction

Health IT (HIT) has become a critical infrastructure in healthcare [1]. The connected use of information-intensive functions (e.g. electronic health records and ePrescribing) has revolutionised the provision of treatment and care. Recently, the HIT landscape has expanded by the use of health apps and social media, empowering patients to take a more active role in their own care [2]. For any technology used in the care pathway, the impact on patient safety is a fundamental concern [3]. HIT has the potential to improve patient safety but also introduce new hazards. For example, ePrescribing can help eliminate transcription errors in a paper-based process but also increase risk by inducing unsafe shortcuts and alert fatigue.

In order to address this challenge, different national reviews have encouraged the healthcare domain to consider and where appropriate adapt practices used in other high-risk sectors, particularly aviation [4], which adopt systematic approaches to safety assurance and management [5]. This typically includes the implementation of a proactive safety management system, generation of a Hazard Log and a safety case and institutionalisation of an open safety culture [6].

In England, the National Health Service (NHS) has been promoting and supporting such approaches for HIT, through a dedicated Clinical Safety Team at NHS Digital. NHS Digital is a public body that is responsible for providing data and IT systems for commissioners, analysts and clinicians in health and social care. Two HIT safety standards, targeting manufactures (SCCI0129 [7]) and health organisations (SCCI0160

[1] Corresponding author: Department of Computer Science, University of York, York, YO10 5GH, UK; E-mail: ibrahim.habli@york.ac.uk

[8]), have been issued by the Standardisation Committee for Care Information on behalf of NHS England. These standards specify normative requirements, supported by informative guidance, for the implementation of a risk management process and demonstration of organisational commitments. Establishing a safety culture has been a primary objective, requiring evidence of commitment by senior management, e.g. providing the necessary resources and a clear chain of responsibility. This includes the appointment of Clinical Safety Officers (CSOs), who, in their capacity as experienced clinicians, are expected to lead the HIT risk management activities.

Similar to the majority of safety processes in other safety-critical sectors such as nuclear [9] and automotive [10], the SCCI0160 and SCCI0129 standards are centred on identifying the hazards posed by the HIT and assessing, mitigating and monitoring the risk associated with these hazards. In this context, a hazard is defined as "*potential source of harm to a patient*" [7], e.g. the patient receives more than the intended drug dose. A clinical risk is defined as the *"combination of the severity of harm to a patient and the likelihood of occurrence of that harm*" [7], e.g. the likelihood that the patient suffers a permanent life-changing incapacity as the result of the drug overdose.

In this paper, through a qualitative case study, we examine current practices in hazard identification, focusing on how clinicians and engineers approach this fundamental task within the HIT safety process, as defined by the SCCI0129 and SCCI0160 standards. We highlight certain technical and organisational challenges and discuss approaches to improving current practices and promoting learning initiatives.

2. Methods

2.1. Setting

This study concerns hazard identification practices for HIT in England, as scoped by the SCCI0129 and SCCI0160 standards. It focuses on the role of hazard identification within the overall risk management process (Figure 1). The standards follow the safety principles established for medical devices and are consistent with ISO14971 [11].

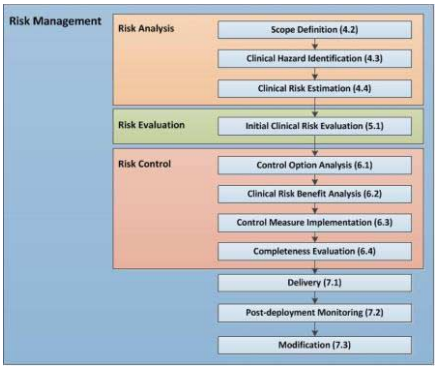


Figure 1. SCCI0129/SCCI0160 Risk Management Activities [7]

Two primary artefacts are generated from the risk management process that explicitly consider HIT hazards: Hazard Log (HL) and Clinical Safety Case Report (CSCR). The HL is a mechanism for recording the on-going identification, analysis and resolution of hazards associated with the HIT system. The CSCR documents an argument,

supported by evidence, for why the system is safe for a given application in a given environment.

2.2. Data Collection and Analysis

Three separate one-day workshops were organised in February/March 2016, involving 34 participants: 19 clinicians, 12 engineers, 2 researchers and 1 patient representative. The participants were selected due to their expertise in the development, deployment and/or assessment of HIT and their understanding of both the engineering and clinical perspectives of the technology. They represented the three main parties involved in HIT risk management: NHS Digital, health organisations and HIT manufacturers. Participants were split into groups of 5. Each group had a moderator who recorded a summary of the discussion. The discussion was led by the following question:

How is Hazard Identification performed so that the hazards identified are specific, relevant, clearly documented and "complete"?

The workshops were followed by detailed reviews of the CSCRs for 20 HIT systems, covering primary and secondly care, based on the above question. The CSCRs considered diverse functions (e.g. care records, prescription, bed management and emergency care) and were submitted by health organisations (for specific deployments), manufacturers (for type approval) and NHS Digital (for the national infrastructure). The CSCR reviews were used to corroborate and augment the workshop outputs.

The data was then imported into NVivo11 for analysis. The text was coded following an iterative process and analysed using Thematic Analysis [12], determining and interpreting repeated patterns of meaning in the data set. The final phase involved combining the different codes into overarching themes using a thematic map, which was independently validated by a senior safety analyst against the original data set.

3. Results

The data indicates that the safety assurance framework established through the SCCI0129 and SCCI0160 standards has provided a systematic approach to identifying hazards within the overall HIT risk management process. When complying with these standards, it is now common practice to produce an explicit HL that is developed by a multidisciplinary team comprising clinicians and engineers. This HL forms the evidence basis for the CSCR. The data also highlighted specific challenges and areas for improvement that concern the technical and organizational aspects of hazard identification. These are summarised in **Table 1** and discussed in the rest of this section.

Firstly, the notion of hazard is not familiar in healthcare settings. The term risk is more recognisable by clinicians, as expressed by one participant: "*the NHS has always worked in the 'risks': don't know what a hazard is*". The overwhelming majority of hazards are care hazards, e.g. patient misidentification, which predate the deployment of HIT and to which the technology now contributes. Positioning the specific hazardous contributions of HIT within the care process is seen as a difficult task.

Secondly, deciding on the level of granularity for hazard identification is problematic. On the one hand, many of the identified HIT hazards are too detailed and correspond to technical failures (i.e. 'network unavailability'). As such, they do not reflect the potential harm to patients. On the other hand, other hazards are defined

generically, with little information about the context, to make them specific to the clinical environment (e.g. 'wrong prescription'). In part, this can be complicated by a poorly-defined clinical scope, as illustrated by one participant: "*an important distinction needed to be made between hazards caused by system and hazards caused by clinical activity. Can the system lead to patient harm or was the patient harm already there but the system perpetuates it?*"

Table 1. HIT Hazard Identification Themes and Recommendations

Summary of Themes
– Confusion about the terms hazard, risk, harm and quality; – Difficulty of positioning hazardous failures of HIT within care processes; – Hazards too detailed to reflect potential harm to patients; – Hazards very generic and poorly linked to clinical environment; – Hazards identified by manufactures lacking validation for their relevance by health organisations; – Lack of early engagement in, funding for, hazard identification; – Perception of hazard identification as a tick-box exercise.
Key Recommendations (made by participants)
– Publish anonymised Hazard Logs for HIT and known hazards of care within the NHS; – Develop practical guidance on hazard identification workshops and techniques; – Develop guidance on the necessary clinical/engineering expertise needed for hazard identification.

Thirdly, it was observed that engineers were more comfortable than clinicians concerning thinking hypothetically about foreseeable hazards, i.e. proactive hazard identification during design stages. Clinicians placed more emphasis on actually experienced hazards and problems based on "*what they already know*". It was noted that many of the events flagged by engineers as hazards were treated as quality issues by clinicians, i.e. events that commonly occur and from which recovery is expected, e.g. 'delay in providing care'.

Fourthly, where do hazards come from? Ideally, the clinicians and engineers from both the manufactures and health organisations should identify the potential hazards collaboratively. A more common scenario has been to take the HL generated by the manufacturers and instantiate it to fit within the specific clinical context of the health organisations. The perception here is that the manufacturers are more competent and have the resources to produce the HL to the required quality. The potential consequence, however, is that many health organisations adopt the HL without the adaptation necessary to cater for the specific local clinical requirements. This is, in part, due to lack of early engagement: "*Poor quality is due to many reasons including doing the work last minute, 'as something that needs to be done', a tick box exercise. It is usually left to the clinician assigned rather than done in plenty of time with a multidisciplinary team. The hazards are generic, often lifted from other documents*". Some highlighted the lack of resources as the primary contributor: "*a continuing message is that there is no funding and resources provided to the NHS to deal with these issues*".

Finally, to increase confidence in the hazard identification results, evidence of the use of systematic techniques is typically provided. What-If Analysis, combined with "*user stories*", appears to be the most common approach. To ensure consistency and promote learning, participants emphasised the need to "*publish anonymised hazard logs for HIT and known hazards of care within the NHS*", combined with "*practical guidance on Hazard Identification workshops and techniques*". Initiatives within NHS Digital are currently focusing on compiling generic hazard logs for different types of HIT systems, including apps, combined with a tool-supported methodology and practical guidance, which will be publicly available for use by the wider community.

4. Discussion and Conclusions

Hazard identification challenges are not uncommon for novel, highly-configurable and context-sensitive technologies. For example, in the automotive industry, autonomous driving has raised concerns about the adequacy of the current approaches to hazard identification, highlighting gaps in our understanding of the relationship between the human driver and the autonomous vehicle functions [13]. For healthcare, ideally, safety analysis should be applied in a top-down and integrated manner, focusing on the potential patient harm and the hazards posed by the health services to which different technologies (including HIT), clinical practices and organisational processes contribute. Such a holistic approach, which is common in aviation and nuclear for instance, is rarely followed in healthcare for many complex reasons [14]. As such, labelling certain HIT failure conditions as hazards is a pragmatic choice and can be criticised as being IT-centric. After all, information, unlike human actions or implantable medical devices, cannot directly lead to harm. However, the sphere of influence for clinicians and engineers who are currently responsible for the development and deployment of HIT is often limited. This has led to treating critical HIT failures as hazards, rather than causes of higher care hazards.

Finally, the recent national review of HIT in the NHS, led by Robert Wachter, highlighted the principle that "*Health IT Entails Both Technical and Adaptive Change*" [15], focusing on clinical aims and practices and patient outcomes (not the mere act of digitisation). Meeting such a principle will help achieve an integrated approach to hazard identification that involves the right clinical and technical stakeholders, including patients and front-line users.

References

[1] S. Agboola, D. Bates, & J. Kvedar. Digital Health and Patient Safety, *JAMA* 315.16 (2016): 16 97-1698.
[2] L.T. Lorchan & J.C. Wyatt, mHealth and Mobile Medical Apps: A Framework to Assess Risk and Promote Safer Use. Journal of medical Internet research 16.9 (2014): e210.
[3] F. Magrabi, S. Ong & E. Coiera, Health IT for Patient Safety and Improving the Safety of Health IT, *Studies in Health Technology and Informatics,* 222 (2015): 25-36.
[4] A.J. Stolzer, M.C.D Halford and M.J.J. Goglia, *Safety Management Systems in Aviation*. Ashgate Publishing, Ltd., 2015.
[5] M. Sujan, I. Habli, T. Kelly, S Pozzi, and C. Johnson, Should Healthcare Providers Do Safety Cases? Lessons from a Cross-industry Review of Safety Case Practices. *Safety Science*, 84 (2016), pp.181-189.
[6] S. Dekker, *Just Culture: Balancing Safety and Accountability*. Ashgate Publishing, Ltd., 2012.
[7] NHS Digital. SCCI0129, Clinical Risk Management: its Application in the Manufacture of HIT Systems. Standardisation Committee for Care Information. 2016.
[8] NHS Digital. SCCI0160, Clinical Risk Management: its Application in the Deployment and Use of Health IT Systems. Standardisation Committee for Care Information. 2016.
[9] IEC, BS. 61513: *2011 Nuclear Power plants-Instrumentation and Control for Systems Important to Safety– General Requirements for Systems*. 2011.
[10] ISO, ISO 26262 *Road Vehicles– Functional Safety*. ISO Standard, 2011.
[11] ISO, ISO 14971: medical devices-application of risk management to medical devices. ISO, 2007.
[12] V. Braun, and V. Clarke, Using Thematic Analysis in Psychology. *Qualitative Research in Psychology*, 3.2 (2006), pp.77-101.
[13] H. Monkhouse, I. Habli and J. McDermid, The Notion of Controllability in an Autonomous Vehicle Context, *Critical Automotive Applications: Robustness & Safety (CARS)*, France, September 2015.
[14] C. Vincent & R. Amalberti, R., *Safer Healthcare. Strategies for the Real World*, The Health Foundation/Springer Open, London, 2016.
[15] R. Wachter, *Making IT Work: Harnessing the Power of Health Information Technology to Improve Care in England*, Department of Health, 2016.

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-559

EHR Improvement Using Incident Reports

Tesfay TEAME[a], Tor STÅLHANE[b], Øystein NYTRØ[b]
[a] *Regional ICT-centre, South East Norway Regional Health Authority*
[b] *Department of Computer Science, Norwegian University of Science and Technology*

Abstract. This paper discusses reactive improvement of clinical software using methods for incident analysis. We used the "Five Whys" method because we had only descriptive data and depended on a domain expert for the analysis. The analysis showed that there are two major root causes for EHR software failure, and that they are related to human and organizational errors. A main identified improvement is allocating more resources to system maintenance and user training.

Keywords: Clinical software; Software Maintenance; Failure analysis; Five Whys

1. Introduction

This paper discusses reactive improvement of Electronic Health Record systems (EHR) viewed as part of a socio-technical system. Proactivity is the ideal for system designers and organizations: Striving to avoid incidents and design flaws instead of fixing or learning not to repeat them. However, systems are more or less flawed, not fulfilling the needs of changing organizations. Thus, our topic is pinpointing and prioritizing improvement in use, design and implementation of an operational EHR system. We describe detected and reported accidents, mishaps, errors or near misses as "incidents", thus not judging or ascribing any reason. "Reactive improvement" means that incidents are analysed to find causes and eliminating them. This way of improving a complex, socio-technical system has been used with success in, e.g., the aviation industry, and it should be used more in healthcare [1]. Many methods support reactive improvement, but we will only employ "Five Whys". Our data consists of error reports from an EHR help-desk in a hospital trust. Note that this work is an initial and independent analysis, not in any way commissioned or expected by the healthcare trust or a software provider. No actions have been taken to inform them about these preliminary results, but we intend to do so in the near future.

The following sections discuss related work, error analysis methods, empirical help-desk data, data analysis, results and finally conclude with recommendations for reactive improvement in practice.

2. Related Work

The EHR fulfils many diverse, complex, partially conflicting requirements. In contrast to non-clinical information systems used in healthcare, the EHR ecosystem provide a very rich context of study. Three areas of research are related to our work: Introduction of IT in healthcare practice and corresponding evaluation of effects in quality, work

practice, patient safety [1]; Sociotechnical aspects of Information Systems [2]; Software and user interface design [3]. Surprisingly little research has focussed on methods for improving design or maintenance of existing systems. It is perhaps revealing that the term "e-iatrogenesis" [4] has been coined to describe adverse effects to the patient because of using clinical software. However, as a means to control and record quality, all specialist healthcare in Norway have implemented systems for recording, reporting and tracking issues related to adverse events [5]. *This paper proposes to take a next evolutionary step: From awareness and systematic recording of deficiencies, causes and effects, to directed improvement and maintenance of systems critical for patient care, i.e. Reactive Improvement.*

3. Reactive Improvement

Reactive improvement is different from incremental, iterative or agile development. 'Reactive' implies that the information system is not in (agile) development or in design, it is considered to be fully deployed, fully operational and in use by the customer organization. The improvement may effect system design, function and use.

What kind of events during operation should trigger improvement? We have chosen to disregard IT-related, e-iatrogenic adverse events, as reported in the separate health quality reporting systems mandated in Norwegian specialist healthcare. This study *only* takes into consideration improvement triggers as reported to the IT helpdesk. In order to make this study useful from a patient and clinical perspective, it could be extended by aligning incidents with related, identifiable health adverse events. However, in this study, we have neither had access to clinical data, nor the analytical methods that would allow us to relate helpdesk event to clinical outcome.

Reactive improvement, based on helpdesk reports, relies on insight into both organization and system design in order to isolate root causes, and propose possible remedies. The EHR interconnects user interface components/devices/subsystems - used in organizational processes - by actors with certain patient responsibilities. In general, we have found it convenient to categorize errors along four independent dimensions:

1. Apparent **situation** of discovery/organizational context
2. **Type** of system malfunction or error
3. Apparent system **location:** a function or module
4. **Seriousness or risk** of patient harm

In the incident reports we have used for this paper, all these dimensions appear. Some, eg. 4, are reported as criticality, while others, eg. 1, must be deducted from the incident report. See incident report sample in section 5. The following sections describe an analysis method, review findings and explain how the data can be applied for ranking and specifying reactive system improvement.

4. Data Analysis Method

We did only have access to textual data and a purpose to extract knowledge, or more precisely, causes for problems with clinical software. Since this is a quite common data analysis problem, meny analysis methods are available. We have chosen "Five Whys" due to its simplicity. The method was first described by T. Ohno [6] and is an

important part of Toyota's production improvement process. Our approach follows Bulsuk [7]. "Five Whys" has been given little research attention, but there exist a large number of blogs reporting practical experiences. Several papers related to health care quality refer to the use of "Five Whys" [8, 9]. The main idea of "Five Whys" is simple – identify the problem, ask experts why the problem did occur and keep on asking "why" five times. The process is as follows:

1. Identify the problem – what are we trying to achieve. Spend some time here. It is important to focus on the root cause and not the symptom.
2. "Why did this happen?" Identify all the causes you can think of.
3. For each of the causes identified in step 2 ask "Why did this happen?"
4. Repeat steps 2 and 3 five times. By this stage, we should have identified all relevant root causes.
5. Identify solutions and countermeasures to the causes identified in steps 2 and 3

The method will run into difficulties if the experts disagree. However, having more than one answer to a why-questions is not a problem. There are two ways to resolve such situations: Either (1) choose one of the alternatives and document why it was chosen or (2) build a tree-structure and select the best alternative later. We can also reason that all remaining root causes eventually will be tackled. Following the five steps described above, the whole process can be documented using a table. When we have reached the last step in the table, we need to suggest actions that will remove or reduce the problem. Without this last step, the whole process is a waste of time.

Subjectivity is the main problem both with "Five Whys" and all other methods of analysis where people try to identify causes. This holds for methods such as Ishikawa diagram – also known as Fishbone diagrams, Fault Trees and Cause – Consequence diagrams.

5. The Data Used

The data used in this paper are collected from a helpdesk system used by Sykehuspartner – the IT service provider owned by the South East Norway Health Regional Authority. All system users in government hospitals of the trust report errors related to clinical ICT to Sykehuspartner, who is responsible for finding causes and solving the problems within an agreed period of time. At the helpdesk, reports are categorized along two dimensions [1], according to criticality (1: threat of life, 2: major/lasting service impact, 3: temporal service impact) and scope/quality (A: Whole wards, sections, patient groups, major loss of efficacy or work effort, B: Smaller groups, C: Individual, but workarounds possible, D: substandard service quality.

This study gathered more than 13000 reports with criticality 1B (i.e. errors that affect a small group of users and that might lead to a situation threatening patient life or being critical to hospital operation) related to different clinical applications, for the period from January 2013 to July 2014. Manual inspection of 13000 reports was not possible, so we narrowed the scope to analysis of 1618 reports related to the EHR. The specific EHR system has wide national coverage, so was of particular interest.

[1] With different responses. Very simplified for the purpose of this presentation

Incidents originate from nine hospitals in the South-East Health Region. Incidents can be have many immediate reasons: system down; malfunction under operation; user error; intra-system communication; data loss; data input failure; retrieval failure etc.

We did a grouping based on the similarity in manifestation of the indicents reported. This was done to avoid redundancy and gain insight into the problem. A representative sample of 26 incidents from each group have been selected and analysed by using the five whys analysis method. A sample incident report from the logging system and the corresponding Five Whys analysis is shown underneath:

Hospital: *YY*
Title: Scheduled contact did not appear in the record window
Description: Scheduled contact does not appear in the medical record registration window. Clinicians cannot register contact diagnosis. This happens intermittently and the clinicians want to know why it happens and how it can be avoided.
Solution: Contacted Dr XX by phone and informed that the clinicians must select/tick the current patient so that the scheduled contact will be visible in the registration window.

Table 2: Example of "Five Whys" analysis

Step	Reason	Why?
1	Clinicians couldn't register diagnosis	Why couldn't clinicians register?
2	Scheduled contact (window) did not appear	Why didn't scheduled contact appear?
3	The clinicians didn't tick the current patient	Why didn't the clinicians tick?
4	The clinicians did not know that ticking was required for registering diagnosis	Why didn't the clinicians know?
5	(New) clinicians didn't get detailed training	Why didn't clinicians get training??
Root cause	***This registration module is not part of the training program for new clinicians.***	

6. Results

In order to classify our resulting problems into categories, we applied the following algorithm:

1. Go through all the root causes and look for frequently used terms
2. Join terms that have the same or close meanings.
3. Repeat step 2 as long as we find categories that may be merged.
4. Select a unifying term for each class

Applying this process to the 26 selected incidents gave the six failure categories shown in table 3 below. Other category sets could have been used but as far as we can see, this would not change the main conclusions.

Table 3: Problem categories

Problem category	Failures	Relative volume
Wrong manual procedures	9	35%
Configuration problems	6	23%
Lack of resources, e.g., for training	5	19%
Deficiencies in (use of) the system monitoring services and brokers	3	11%
Unknown cause	2	7%
Long delays in a specific external register	1	5%
Sum	26	100%

We see that the main problem categories are not technical, but related to human users (both clinical and IT) and their organizations. Wrong manual procedures, configuration problems and lack of resources for training and monitoring.

7. Conclusions and Implications

One broad conclusion that can be drawn based on our results is that Norwegian specialist healthcare may be unprepared for the challenges of continuous implementation of new information technology. Two categories of reasons stand out:

1. Too few resources are allocated for training the users and for using and responding to issues related to the service control/monitoring system.
2. Manual procedures are not well tested and validated.

Both of these problems are solvable, but require that the hospital administration understand them and give them priority. An additional challenge is that inferior log data quality makes many reported incidents impossible to analyse. The hospitals and their staff waste important improvement opportunities. In our opinion, the incident data quality will improve when the persons involved see that it is used for something important – namely to improve their systems and the service that they provide. On a higher level, in order to improve the quality of service offered to users and patients, it is necessary, but challenging, to analyse the relationship between EHR incidents and health outcomes. EHR malfunction does impact outcome [11], sometimes by improving quality at the cost of lowering efficiency. Our perspective on reactive improvement hunts the "why" of causality in the IT system. Hunting "whys" in the clinical work process is beyond our methods, but should nevertheless have high priority.

Both hospitals and software development companies should look at all problems and failure reports as an opportunity to improve their processes. This requires good reporting, an open communication and the necessary resources. All problems should be analysed using e.g. "Five Whys" and the identified root causes should be understood, resolved or at least remedied so as not increase patient risk.

References

[1] J. Talmon et al.: *STARE-HI – Statement on reporting of evaluation studies in Health Informatics*. International journal of medical informatics, vol. 78, number 1, 2009
[2] C. Bossen et al.: *Data-work in health care: The New Work Ecologies on Healthcare Infrastructures.* Proc. 19th ACM Conf. on CSCW, CSCW-16, 2016
[3] B. Middelton et al.: *Enhancing patient safety and quality of care by improving the usability of electronic health record systems*. J AM Med Inform Assoc, Jun; 20(e1): e2-e8, 2013.
[4] JP. Weiner et al.: *"e-Iatrogenesis": The Most Critical Unintended Consequence of CPOE and other HIT*. JAMIA. 2007; 14(3).
[5] Law on specialist health care services §3-3a, §3-4a, https://lovdata.no/dokument/NL/lov/1999-07-02
[6] T Ohno: *Toyota production system: beyond large-scale production*. CRC Press, 1988.
[7] K. Bulsuk: http://www.bulsuk.com/2009/03/5-why-finding-root-causes.html
[8] M. Daly: *Healthcare Audit Criteria and Guidance Prepared by the Clinical Audit Criteria and Guidance Working group*, Health Service Executive, Ireland, Aug. 2008.
[9] P. Spath: *Introduction to Healthcare Quality Management*, Health Educations Press, 2009.
[10] P. Ladkin and K. Loer: *Why-Because Analysis: Formal Reasoning About Incidents,* RVS-Bk-98-01, Technischen Fakultät der Univ. Bielefeld, Germany, 1998.
[11] Wang Y, Coiera E. et al.: *Measuring the effects of computer downtime on hospital pathology processes*. *Journal of Biomedical Informatics*. 2016; 59:308-15.

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-564

Improving Handovers Between Hospitals and Primary Care: Implementation of E-Messages and the Importance of Training

Grete NETTELAND[1]
Sogn og Fjordane University College, Norway

Abstract. The transfer of information and responsibility for care of a patient from one healthcare provider to another is referred to as a handover. While some handovers are effective and achieve high quality communication, others represent a barrier to continuity of care. To increase the patient safety, Norway decided to replace handovers with an *electronic e-message system (EMS)*. This paper refers to a quantitative study of this implementation and examines the opinions of first-line leaders and nurses (N=108) on how organisational factors were taken into account and how the implementation might be improved. The findings indicate that such factors generally did not receive very much attention in the implementation of the EMS, and less for the nurses than for the first-line leaders. Particularly, the factor most prominently identified by both groups as warranted improvement, was the training.

Keywords. E-messaging, handover, implementation, training, organizational factors.

1. Introduction

Handovers are described as transfer of responsibility and accountability for patient care from one provider or team of providers to another [1]. They occur within and between organisations, and the communication takes place either face-to-face, by telephone, e-mail, fax or electronic messaging. The desired outcome is patient safety and a reduced number of failures. But handovers may also represent a barrier, or even worse, be a high-risk point in patient care [2, 3, 4]. Patient information might get lost, ignored or misinterpreted, treatment might be delayed and patients may die. In 2008, Norwegian health authorities, in line with national strategies, decided to develop an electronic-messaging system (EMS) that could make selected patient information available for all partners in the health and care sector, provide more efficient work and a safer continuity of care. Five years later, a set of standardised e-messages with content customized for the transition between hospitals and municipal healthcare was disseminated nationally [5]. The role of the nurses was to produce adequate and correct information in e-messages, transfer these messages to the next health provider, and, to control and read the content in the e-messages they received.

Despite this, there are still challenges in the handovers between hospitals and municipal healthcare. These are so far explained by 1) the lack of integration of e-messages into day-to-day work; 2) a general lack of ICT skills among the healthcare staff; 3) the functionality of the EMS; 4) the system's usefulness for different user groups; and, 5) the varying quality of the e-message content, sometimes being incorrect,

[1] Corresponding Author: Grete Netteland, Dr. Polit. SFUC. E-mail: grete.netteland@hisf.no

incomplete, inconsistent or delayed [6, 7, 8]. It is further recognized that the perceived challenges might differ between nurses and FLLs and that e-messaging might affect the collaboration between different groups regarding tasks and responsibilities [3, 9]. Implementations of e-health systems have been assessed as demanding. The same seems to apply to the implementation of e-messaging. This paper will examine four organisational factors frequently mentioned in systematic reviews as being critical to ICT implementations in the healthcare sector [4, 10]: *information* (e.g. sense-making and the announcement of the project's importance, scope, progress, objectives and activities [2, 4, 11]); *involvement* (e.g. individual and collective involvement in planning and implementation, collective action and interaction [4, 10, 11]); *training* (e.g. adequate training of staff members and consistent use of the new system [4, 11]), and, *support* (e.g. skilled people that can assist when needed and follow-up from manager [8, 10]).

Because the e-message system is a new tool for handovers and interorganisational collaboration, this study is inspired by a sociotechnical perspective [12]. With this perspective as a departure point, the implementation and integration of the e-message system into different organisations will depend on the interaction among individuals and organisational as well as social and technological elements in local settings. This paper addresses how the organisations have prepared their nurses and FLLs for the organisational factors mentioned above, in implementing the e-messages, and how the implementation process might be improved.

2. Method

The study is based on an online survey conducted in a Norwegian county in spring 2014. Data were collected from 1) the nursing homes and home healthcare services in three municipalities selected as representatives of the three EHR systems in use in the county, and, 2) the three units in the main hospital with the highest rate of e-message exchange with municipalities. The survey took place four to five months after the e-message implementation in the respective units. Three sources formed the basis for the development of the questionnaire: Interviews with the project teams in the county hospital and in one municipality; careful examination of the national e-message program directive, national and regional implementation guidelines and national pilot reports [5, 6, 13]; and, research on handovers, implementations of e-health systems, and the introduction of e-messages in different contexts. Before the questionnaire was launched, it was sent for comments to two FLLs and two project leaders, asking them to involve one or two nurses. The final questionnaire, which had 35 questions for nurses and 41 (6 extra) for FLLs, included the following topics: demographics, ethics, organisational and implementational characteristics, experienced challenges, deviations and errors, perception of the implementation and the EMS, and suggestions for improvements. It was distributed via a link to nurses and leaders who were on duty in their respective units in a 24-hour period. The response rate was 93% for FLLs and 90% for nurses. Totally 93 nurses (4male) and 15 FLLs (1 male) responded. The responses to most of the questions involved rating on a six-point Likert scale (1 = to a very small extent; 6 = to a very large extent). Some questions permitted a binary choice in the response and a few required free responses. This paper is based on twelve items focusing on how the implementation had prepared the nurses and the FLLs for information, involvement, training and support; two items focusing on the participants' perception of how the national aims more efficient work and a safer continuity of care had been achieved; and, one open-ended

question dealing with how to improve the implementation. The latter responses were coded according to the items and organisational factors before the meaning was condensed [14]. All analyses were carried out using IBM SPSS ver. 23. The t-tests were carried out by conducting Student's t-test, which has a threshold of $p \leq 0.05$.

3. Results – Assessment of Organisational Factors

The main focus here is on 1) the descriptive differences in the organizational implementation data rated using a Likert scale, and 2) the responses from the open-ended question. The first results are split into two separate tables, both listing the mean values and standard deviations of the ratings for FLLs and nurses. For all variables the differences in mean values between FLLs and nurses are statistically significant ($p \leq 0.05$). While Table 1 describes variables related to the themes of *information* (the five columns to the left) and *involvement* (the remaining columns). Table 2 describes variables related to *training* (the four columns to the left) and *support* (the last column). Clear differences in mean values between FLLs and nurses are found for all variables in both tables, the clearest ones in Table 1 for variables characterised as *involvement* (2.15 and 2.07). Rather large differences of this kind, albeit smaller, were identified in the same table for *information-related* variables, where differences in mean values between the two groups ranged from 1.82 to 1.17, respectively. It should be noted that the mean values in Table 1 were higher for the FLLs (4.80 to 4.27) than for the nurses (3.63 to 2.12). The values of the nurse responses are more widely spread than the values of the FLL responses. Overall, Table 1 indicates that the FLLs considered that they were better informed about the project and more involved in the planning.

Table 1. Information and involvement – nurses (N) and first-line leaders (FLLs)

	Good information about the project		Necessary information about progress		Information about project goals		Information about success criteria		Introduced to ethical and security-related issues		Involved in the planning		Opportunity to influence the introduction	
	N	FLLs	N	FLLs	N	FLLs	N	FLLs	N	FLLs	N	FLLs	N	FLLs
Mean	3.63	4.80	3.62	4.80	3.53	4.73	3.14	4.73	2.71	4,53	2.26	4.33	2.12	4.27
Standard dev.	1.420	1.082	1.414	1.320	1.515	1.100	1.434	.884	1.571	.743	1.444	1.676	1.451	1.387

Table 2 focuses on the assessment of training and support. As in Table 1, the assessment ratings differ between the two groups, with a maximum difference between mean values of 1.51 for the variable *Sufficient time for testing*. The smallest difference in mean values is found for the variable *Relevant training* (1.04). Within each of the user groups, the mean values for training and support variables range from 3.69 to 2.89 for nurses and from 4.73 to 4.40 for FLLs. Compared with Table 1, the assessment across variables within each group is smoother. The spread within the different variables is also smaller. Both for nurses and FLLs, the lowest mean value was related to the variable *Sufficient time for testing* (2.89 and 4.40 respectively). On the other hand, the absolutely highest mean value for nurses was for the variable *Sufficient training* (3.5), for leaders the variable *Relevant training* (4.73). Considering all of the participants, Table 2 suggests that the leaders as a whole probably found the given training more adequate and more relevant than the nurses.

Table 2. Training and support- nurses (N) and first-line leaders (FLLs).

	Received sufficient training		Learnt to report mistakes and deviations		Relevant training		Sufficient time for testing		Sufficient support	
	N	FLLs	N	FLLs	N	FLLs	N	FLLs	N	FLLs
Mean	3.41	4.53	3.11	4.60	3.69	4.73	2.89	4.40	3.34	4.60
Standard dev.	1.446	1.187	1.710	.910	1.367	1.033	1.543	1.298	1.514	1.183

In the open-ended questions both user groups provided comments on the 12 variables (Table 3). Eight of the fifteen FLLs suggested improvements to the training, three of them to the information, three to the involvement and six to the support. The respective allocations of these factors among nurses were respectively 58, 22, 1 and 22. The factor most prominently identified by both groups as warranting improvement was the training especially in the nurse group where 62% of the nurses answered that more adequate and different training would improve the implementation. It should be noted that there was a significant medium correlation between *Training* (an aggregate of the four training-related variables) and *More efficient work* (r=0.475) as well as between *Training* and *Safer continuity of care* (r=0.419), assessed according to Cohen's rules of thumbs [15]. This will be addressed in later work. The nurses' attention to involvement was minimal.

Table 3. Distribution of the four organisational factors in open-ended questions - nurses (N) and first-line leaders (FLLs)

Respondents	n=	Information		Involvement		Training		Support	
N	93	22	24%	1	1%	58	62%	22	24%
FLLs	15	3	20%	3	20%	8	53%	6	40%

4. Discussion

Two particularly striking findings were made in this study, namely, the differences in mean values between how the FLLs and nurses assessed the organisational factors and the substantial need for training. The most marked differences were in the mean values for the variables addressing involvement and information, and some smaller, but still evident, for training- and support-related variables. This is interesting as it shows that, while the leaders are generally well informed about project progress, success criteria and so on, the nurses, who are expected to use the e-messages as part of their daily work, to a large extent lack the same information. Together with the low mean value of the nurse involvement, this might indicate that information and anchoring had stopped at the leader level, or, that the organisation had underestimated the needs in the nurse group. This happened although national and regional project guidelines underlined the need for organizational changes and suggested to arrange for satisfying training activities and user support adjusted to relevant target groups [6, 13]. At this stage, it remains unclear whether the implementation for the FLLs was better planned, or that this group received more information, training and support, or was more involved in the implementation process. Another explanation of the differences of opinion is that the two groups differ in their needs for training and information, among others, given their different occupational roles, as claimed by McGinn et al. [10]. It is also interesting that, despite mean values for training and support for nurses lying in the middle of the scale and being higher than the corresponding levels for information and involvement, the poor quality of the training was the main focus of comments in the open-ended responses. Some examples: more

individual and collective training before, during and after the implementation, more information about how to use the different e-messages, and the lack of opportunities for real time testing across collaborating organisations. This detailed description of required types of training, gives a valuable contribution to future training programmes, since these issues often are discussed superficially in the literature [11]. The identified correlations between training and the two national aims, more efficient work and a safer continuity of care, are also interesting. Suggestions for improvement also came from leaders, but these were less detailed about exactly what should be done. A few of the nurses also requested higher information quality in the received e-messages, which supports Bjørlo et al.' finding [8]. Overall, the low mean values for the organisational variables in the nurse group and the numerous requests and suggestions for improvement in this study, might indicate that there is plenty of room for improvement in the implementation of e-messaging.

5. Conclusion

The findings indicate that organizational factors and in particular training should receive more attention when e-messages are implemented across administrative levels. For future implementations[2], a more detailed and targeted training programme should be developed taking into account 1) that nurses and FLLs might assess the organisational factors differently, and, 2) that actors in collaborating organizations should be involved.

References

[1] W. Chaboyer, *Clinical Handover*, Griffith University, 2011.
[2] N. Staggers, J.W. Clark, Research on nursing handoffs for medical and surgical settings: an integrative review, *Journal of Advanced Nursing*, **69** (2) (2013), 247-262.
[3] L. Melby, R. Hellesø. Introducing electronic messaging in Norwegian Healthcare: Unintended consequences for inter-professional collaboration, *IJMI* **83** (2014), 343-353.
[4] F. Chang, N. Gupta, Progress in electronic medical record adoption in Canada. *CFP* **61** (2015), 1076-1084.
[5] The Norwegian Directorate of Health, *National initiative of e-messaging: Final report*, Oslo, 2012.
[6] T.S. Bergmo, G. Ersdal, E. Rødseth, G. Berntsen, Electronic Messaging to Improve Information Exchange in Primary Care. *TELEMED 2013* (2013), 172-177.
[7] M. Lyngstad, D. Hofoss, A. Grimsmo, R. Hellesø, Predictors for Assessing Electronic Messaging between Nurses and General Practitioners as a Useful Tool for Communication in Home health Care Services: A Cross-Sectional Study, *Journal of medical Internet research*, **17**(2) (2015): e47.
[8] A.R.B. Bjørlo, H. Christensen, R. Fensli, Electronic messaging- a contribution to fulfil the Coordination reforms intentions of coherent, seamless, coordinated and safe health services? *SHI* (2014), 105-106.
[9] G. Netteland, Exploring implementations of electronic nurse and care messaging at municipality level, *SHI* (2015), 102-105.
[10] C.A. McGinn, S. Grenier, J. Duplanti, N. Shaw, C. Sicotte, L. Mathieu, Y. Leduc, F. Légare, MP. Gagnon, Comparison of user groups' perspectives of barriers and facilitators to implementing electronic health records: a systematic review, *BMC Medicine* (2011), 9-46.
[11] F.S. Mair, C. May, C. O'Donnell, T. Finch, F. Sullivan, E. Murray, Factors that promote or inhibit the implementation of e-health systems: an explanatory systematic review, *BWHO* **90** (2012), 357-364.
[12] M. Berg, J. Aarts, J.van der Lei J, ICT in health care: sociotechnical approaches, *MIM* **42**(4) (2003), 297-301.
[13] Bergen kommune, *Vestlandsheftet: Rettleiar for utbreiing av elektroniske meldingar,* 2013.
[14] K. Ringdal, *Research and Quantitative methods in Social Sciences*, Fagbokforlaget, Bergen, 2012.
[15] J. Cohen, *Statistical power analysis for the behavioral sciences*, Routledge, 2013.

[2] It should be noted that the two FLLs who commented on the questionnaire, took part in the study.

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-569

The Association Between the STOPP/START Criteria and Gastro-Intestinal Track Bleedings in Elderly Patients

Anouk VELDHUIS[a], Danielle SENT[a,1], Linette BRUIN-HUISMAN[b], Erna BEERS[b], Ameen ABU-HANNA[a]
[a] *Department of Medical Informatics, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands*
[b] *Department of Family Medicine, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands*

Abstract. Potentially inappropriate prescribing is a common problem, especially in elderly care. To tackle this problem, Irish medical experts have developed a list of criteria when medication should be added or omitted based upon the patient's physical condition and medication use, known as the STOPP and START criteria. The STOPP and START criteria have been formulated to identify the prescribing of potentially inappropriate medicines (PIMs) and potential prescribing omissions (PPOs). One of the most common problems of inappropriate prescribing is gastro-intestinal track bleedings. For this purpose, nine of the 87 STOPP and START criteria are designed to prevent this. However, the prevalence of gastro-intestinal track bleedings has not been established when these nine STOPP and START criteria are violated. The database contained 182,000 patients belonging to 49 general practitioners in the region of Amsterdam, The Netherlands. We estimated both the incidence of PIMs and PPOs and whether harm, in this case a gastro-intestinal track bleeding, occurred. We found that although violation of the nine STOPP or START criteria were possibly associated with harm (OR = 1.30), this association was not statistically significant (p = 0.323). Searching for evidence for harm informs decision support design aimed at improving quality of medication prescription as it prioritizes the many suggested criteria based on their relevance.

Keywords. STOPP START criteria, gastro-intestinal track bleeds, elderly people, primary care, CPOE, NSAID, adverse drug reactions, drug safety, older patients.

1. Introduction

Life expectancy increases as the average life standards and medical developments improve over time. For example, in the Netherlands, the life expectancy at birth in 1995 was 77.4 years whereas it was 81.9 years in 2015 [1]. Although medications play an important part in prolonging life expectancy, they are also associated with risks of

[1] Corresponding author, D. Sent, Dept. of Medical Informatics, University of Amsterdam, PO Box 22700, 1100 DE Amsterdam, The Netherlands; E-mail: d.sent@amc.uva.nl

adverse events. This is especially relevant in elderly care where the contribution of medications to health utility is not well understood.

In 1991, Beers et al. developed criteria to look into the possible prescribing problems in elderly patients to prevent harm [2]. As this list was based on the American health care system, a group of Irish experts developed a new list of criteria, the STOPP and START criteria, to adjust and expand Beers' list for the European situation [3]. The STOPP criteria are concerned with medications that should be stopped, the potentially inappropriate medicines (PIMs);. The START criteria concern medications that are currently omitted and should be started, the potential prescribing omissions (PPOs). Both categories of criteria are based upon the patient's physical condition and use of other medications. The original list consists of 65 STOPP and 22 START criteria. It was later translated and adjusted to suit the Dutch situation [4,5].

The Beers criteria and the STOPP and START criteria are both based on general pharmacological knowledge [2,3]. In general, there is paucity in studies investigating the empirical harm of individual STOPP and START criteria. This is problematic since it is hard to focus on the most relevant criteria from the large number of the STOPP and START criteria. The main objective of this study is to gain empirical evidence for the STOPP and START criteria associated with the important side effect (harm) of gastro-intestinal bleeding. The (automatic) identification of PIMs and PPOs, for instance by using a decision support system (DSS), can contribute to optimize strategies and prevent the implied harm pertaining to the START and STOPP criteria.

This paper is organised as follows: in Section 2 we will describe the decision tool and the conducted study. In Section 3 we will represent the results, followed by a discussion in Section 4. We will end this paper with our conclusions in Section 5.

2. Methods

Our study forms a secondary analysis of data originally collected to investigate the incidence of PIMs and PPOs by Bruin-Huisman et al [6]. The database of the Academic Medical Center of Amsterdam, the Netherlands, contains 182,000 patients of 49 general practitioners (GPs). Our sample consists of all patients aged $\geq$ 65 years on January 1st of the investigated years (2007-2014) with $\geq$ 1 prior year of data and $\geq$ 1 record in the investigated year or in the subsequent six months.

Selection of criteria pertaining to gastro-intestinal bleedings was performed on both the original STOPP and START list of Gallagher et al. as well as the Dutch list [3,4]. The six criteria selected from the original list (A9, A11, A12, A17, E1 and E5) were matched with their equivalents on the Dutch list. Furthermore, both lists were checked for other criteria, which could logically lead to the development of gastro-intestinal bleedings. This resulted in two additional criteria from the Dutch list being added to our selection (D2A and E5). The complete list of criteria is STOPP A9, A15 and E5, and START D2A, D2B, D2C, D3A, D3B and D3C. Criteria A9, D3A, D3B and D3C involve the use of acetylsalicylic acid or carbasalate calcium (blood thinners). Criteria E5, D2A and D2B involve nonsteroidal anti-inflammatory drugs (NSAIDs). Criterion D2C involves both. Most criteria discuss the absence of a PPI (protects the gastric wall). Criterion A15 is about the use of platelet inhibitors (blood thinners) in patients with a concurrent bleeding disorder.

To determine whether harm in the form of gastro-intestinal bleeding had occurred, two selection methods were used. The first is based on the International Classification

of Primary Care (ICPC) codes, used by the GP to document diagnoses. All codes describing gastro-intestinal bleeding, or its symptoms, were included (Table 1). The second method consisted of screening the health record for keyword combinations, which could be associated with gastro-intestinal track bleedings. To combine keywords, wildcards for partial matching were used. We considered a case as indicating harm if either a selected ICPC code was used or one of our keyword combinations was found. We considered harm due to (violating) a STOPP or START criterion if it occurred between the prescription date and the end of the prescription date plus fourteen days. The latter date was computed using the prescription date, the number of pills prescribed and the number of pills prescribed per day. Fourteen days are added to compensate for lag time between the prescription and pharmacy pick up date, and for missed doses, which result in a longer therapy duration.

Table 1. Selection criteria for Gastro-intestinal track bleedings

ICPC	Diagnosis
D14	Haematemesis
D15	Melaena
D85	Ulcus duodeni
D86	Ander ulcus pepticum (ulcus ventriculli)
keywords	"Bloed%" (Blood) and: "Maag%" (Stomach) "Digest%" "Tr Dig%" "Gastro%"

Patient records have been analyzed for the years 2007-2014. We reused the selection method used in Bruin-Huisman et al [6]. For each year we determined the total harm and harm potentially due to a PIM or PPO. Furthermore, we calculated an odds ratio on harm for patients with one of the selected STOPP or START violations in comparison to patients who did not have a STOPP or START for each year separately.

3. Results

In Figure 1, the number of PIMs and PPOs during harm is noted as 'event'. The odds ratio varies between 0.21 (when only 1 patient could be identified) and 2.42. Of the 8 investigated years, 4 years indicate an odds ratio greater than 1 and 4 years a value below 1. The pooled odds ratio for all years combined is 1.30 (p=0.323).

Table 2 shows the number of gastro-intestinal track bleedings per STOPP and START criterion per year and overall. Five criteria show a number of PIMs or PPOs that vary greatly. Harm given D2A or D2C is registered one time in eight years, while harm given D3C has been documented 14 times in the eight studied years. From 2007 to 2012, harm only occurred during the last 3 criteria, D3A, D3B and D3C. No PIMs were found for patients who suffered from a gastro-intestinal bleeding at the same time.

4. Discussion

A positive association between violation of a selected START or STOPP criterion and the development of gastro-intestinal track bleedings was not proven, as the pooled OR showed a small positive, but not significant, association.

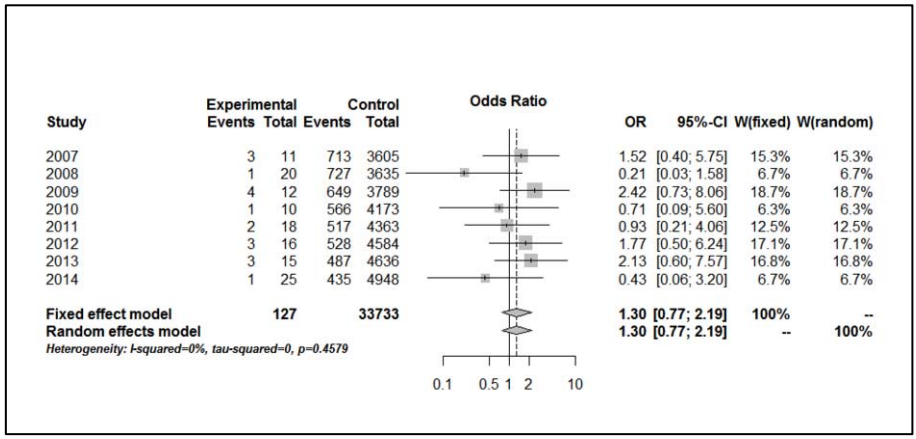


Figure 1. The odds ratio's for the years 2007-2014 and a pooled odds ratio for all years combined.

Additionally, the prevalence of harm, as we defined it within the STOPP or START period is very small. Of the nine STOPP and START criteria designed to prevent gastro-intestinal bleeding only five showed actual harm possibly due to violating the criteria.

This study is the first to focus on the implications in practice of the STOPP and START criteria. By doing so, an example can be made to understand the relevance for such a list for GPs in daily practice. We had a large dataset which we could use to assess and select relevant data. Furthermore, we made an analysis of harm, which included multiple methods, thereby taking into account bias due to the possibility of suboptimal registration of ICPC codes and the differences in describing certain diseases between GPs.

As we reported harm due to PIMs and PPOs as occurring only during and immediately after the prescription period, a larger percentage of harm might be associated with these PIMs and PPOs. The harm as a consequence of possibly incorrect prescribing behavior could occur after the STOPP or START period is finished. A different analysis could have shown this effect. The reported odds ratio would then increase.

Table 2. The number of patients with harm within the STOPP or START period.

year	harm A9	harm A15	harm E5	harm D2A	harm D2B	harm D2C	harm D3A	harm D3B	harm D3C
2007	0	0	0	0	0	0	1	1	2
2008	0	0	0	0	0	0	1	0	0
2009	0	0	0	0	0	0	2	1	3
2010	0	0	0	0	0	0	0	0	1
2011	0	0	0	0	0	0	1	0	2
2012	0	0	0	0	0	0	1	1	3
2013	0	0	0	1	0	1	1	0	2
2014	0	0	0	0	0	0	0	0	1
all yrs	0	0	0	1	0	1	7	3	14

Another limitation of our study is the lack of data on recipe-free medications, such as certain NSAIDs. As NSAIDs are part of criteria E5, D2A, D2B and D2C, but will not be registered as such if they are not prescribed by a GP, under-registration of these three criteria could occur. In some cases, no PIMs or PPOs were found during a year. Any harm during that year cannot be a consequence of a STOPP or START. To eliminate this problem, a larger group of patients is required or research to provide insight into the necessity of a certain STOPP or START criterion.

Further research could focus on the impact of the other major side-effect groups that are covered by other STOPP and START criteria than the ones used in this study. It might be possible to edit the original STOPP and START list to develop a list consisting of only criteria that have supporting evidence pertaining to their relevance. These relevant criteria can then be used to build a focused DSS (CPOE), where advice is given based on known harmful medications or combinations. By only using relevant criteria, alert fatigue is less likely to occur and offered advice is easily followed. This is the strength of our study as a predecessor to the development process.

5. Conclusion

We analyzed a database of 182,000 GP patients over a period of 8 years to investigate whether or not harm is associated with preventable inappropriate use of medication. Although a possible association was found between potentially inappropriate medicines (PIMs) and potential prescribing omissions (PPOs) and gastro-intestinal track bleedings, this observed correlation was not significant. Our results show a large difference in detected harm of the different PIMs and PPOs related to gastro-intestinal track bleedings. If this difference would be confirmed, the current list of STOPP and START criteria could be adapted and improved. When implementing STOPP and START criteria in decision support systems (such as a CPOE), one should focus on the relevant criteria in order to prevent alert fatigue, but still improve healthcare by preventing possible harm.

References

[1] Global Health Observatory data repository, World Health Organization (WHO). 2016. World health statistics 2016: monitoring health for the SDGs. Available from: http://www.who.int/gho/publications/world_health_statistics/2016/en/

[2] M.H. Beers, J.G. Ouslander, I. Rollingher, D.B. Reuben, J. Brooks, J.C. Beck, Explicit criteria for determining inappropriate medication use in nursing home residents, *Archives of Internal Medicine*, **151(9)** (1991), 1825–32.

[3] P. Gallagher, C. Ryan, S. Byrne, J. Kennedy, D. O'Mahony, STOPP (Screening Tool of Older Persons' Prescriptions) and START(Screening Tool to Alert doctors to Right Treatment).Consensus validation, *International Journal of Clinical Pharmacology and Therapeutics*, **46(2)** (2008), 72–83.

[4] M. Verduijn, A. Leenderts, A. Moeselaar, N. De Wit, R. Van Marum, *Multidisciplinaire richtlijn Polyfarmacie bij ouderen* (Multidisciplinary guideline: Polypharmacy for elderly patients), NHG (Dutch Association of General Practisioners), 2012.

[5] A.M.A. Vermeulen Windsant-van den Tweel, M.M. Verduijn, H.J. Derijks,R.J. van Marum, Stand van zaken Detectie van ongeschikt medicatiegebruik bij ouderen Worden de STOPP- en START-criteria de nieuwe standaard? *Nederlands Tijdschrift voor geneeskunde*, **156** (2012), A5076

[6] L. Bruin-Huisman; A. Abu Hanna; H. van Weert, E. Beers, Potentially Inappropriate Prescribing to Older Patients in Primary Care: A Retrospective Longitudinal Study, *Age and Ageing* (2017)

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-574

Needles in a Haystack: Screening and Healthcare System Evidence for Homelessness

Jamison D. FARGO[a,1], Ann Elizabeth MONTGOMERY[b], Thomas BYRNE[c], Emily BRIGNONE[a], Meagan CUSACK[d], Adi V. GUNDLAPALLI[e]
[a]*VA Salt Lake City Health Care System and Utah State University, Logan, UT, USA*
[b]*VA Birmingham Health Care System and University of Alabama at Birmingham, Birmingham, AL, USA*
[c]*VA Boston Health Care System and Boston University, Boston, MA, USA*
[d]*Corporal Michael J. Crescenz VA Medical Center, Philadelphia, PA, USA*
[e]*VA Salt Lake City Health Care System and University of Utah, SLC, UT, USA*

Abstract. Effectiveness of screening for homelessness in a large healthcare system was evaluated in terms of successfully referring and connecting patients with appropriate prevention or intervention services. Screening and healthcare services data from nearly 6 million U.S. military veterans were analyzed. Veterans either screened positive for current or risk of housing instability, or negative for both. Current living situation was used to validate results of screening. Administrative evidence for homelessness-related services was significantly higher among positive-screen veterans who accepted a referral for services compared to those who declined. Screening for current or risk of homelessness led to earlier identification, which led to earlier and more extensive service engagement.

Keywords. Homeless, Screening, Veterans

1. Introduction

Early identification of homelessness is imperative for successful prevention or intervention efforts, which can include providing appropriate services or directing individuals to existing resources. Apart from homeless shelters and other similar agencies, healthcare clinics, hospitals, and systems can serve as a means to identify individuals who are experiencing homelessness or risk. Although screening for homelessness has been attempted in a variety of settings, empirical evidence of its efficacy is lacking. Accurately targeting individuals with the greatest need for homelessness prevention and intervention—and forecasting the onset of homelessness based on known risk factors—is complicated and often associated with a high false positive rate, making homelessness prevention relatively inefficient [1-2].

A population that is particularly vulnerable to homelessness is U.S. military veterans, who, as compared to the general population, are over-represented among individuals experiencing homelessness. Despite significant policy and programmatic steps in the U.S. Department of Veterans Affairs' (VA) shift toward homelessness prevention, limited evidence exists as to which prevention-oriented strategies implemented by VA mitigate homelessness risk. Improving measurement of

[1] Jamison D. Fargo, 2810 Old Main Hill, Logan, UT, 84322, USA; E-mail: jamison.fargo@usu.edu.

homelessness risk, identifying veterans at greatest risk, and intervening effectively is essential to ongoing VA initiatives.

In 2012, the VA deployed a largescale implementation of a screening instrument known as the Homelessness Screening Clinical Reminder (HSCR), which presents a unique opportunity to quantify the prevalence of homelessness [3-4]. Additionally, limited research on homelessness risk has examined approaches to engage persons identified as at imminent risk for homelessness and provide services to mitigate such risk [5]. Rigorously examining the validity and efficacy of this instrument will ensure the effective use of limited resources. The HSCR is designed to link veterans immediately with Veterans Health Administration (VHA) homeless programs or social work services, but its effectiveness at making these linkages is unclear. Therefore, the objective of this evaluation study was to evaluate criterion validity of the HSCR, as well as the processes by which veterans who are currently experiencing homelessness or risk are linked with services.

2. Methods

2.1. Homelessness Screening Clinical Reminder

To assist in the identification of veterans in need of homelessness prevention, the National Center on Homelessness Among Veterans in Philadelphia, PA, USA, developed the HSCR, a two-question universal screener that assesses housing instability and risk among veterans who present for outpatient care and are not already engaged with VHA homeless programs. The two questions are:

- *In the past 2 months, have you been living in stable housing that you own, rent, or stay in as part of a household?* ["No" indicates veteran is positive for current housing instability]

- *Are you worried or concerned that in the next 2 months you may NOT have stable housing that you own, rent, or stay in as part of a household?* ["Yes" indicates veteran is positive for risk of housing instability]

Veterans who screen positive to either question are then asked where they have lived for most of the previous two months and whether they want to be referred for services. In conjunction with the veteran's self-reported living situation, results of screening were used to form three major groups [6]: 1) positive screen for housing instability, 2) positive screen for risk of housing instability, 3) negative screen for housing instability or risk; and then group 1 was further subdivided into two groups: 1a) positive screen for housing instability with a current homeless living situation (i.e., shelter, on the street, with a family member or friend [doubled-up], or in a motel/hotel) and 1b) positive screen for housing instability without a current homeless living situation (i.e., subsidized or unsubsidized housing, or in an institution). Group 1a represented the most stringent classification for current housing instability or homelessness. Veterans whose current living situation was "Other" at the time of screening were not included in groups 1a or 1b.

2.2. Data

Veterans Informatics and Computing Infrastructure (VINCI) provides secure access to VA data sources through an integrated suite of databases in a secure, high-performance-computing environment [7]. VINCI houses data on over 21 million veterans nationwide. Data available on VINCI for this study included veterans' demographic, military, and healthcare characteristics (International Classification of Diseases, Ninth Revision, Clinical Modification [ICD-9-CM] codes, clinic stop codes, inpatient treatment specialty codes, and National Homeless Registry data [longitudinal data for veterans who have experienced homelessness]). Homelessness was defined as receiving a clinical diagnostic code (ICD-9 v60.0) indicative of housing instability, participation in a VA homelessness-related clinical service or treatment specialty, and/or participation in a VHA homeless program within 90 days of initial HSCR response [8]. Results of the HSCR were matched with demographic, military service, and healthcare data.

2.3. Sample

A total of 5,845,937 veterans were asked to complete the HSCR between October 1, 2012 through September 30, 2014. However, 74,441 (1.3%) were excluded or not screened because they 1) reported that they were already receiving housing assistance (n=11,020), 2) declined screening (n=2,656), 3) were a nursing home resident (n=1,202), 4) were unable to perform screening (n=148), 5) used a VHA homeless program in the 6 months prior to screening (n=57,356), or 6) were missing screening results (n=2,059). This resulted in a final sample of 5,771,496 veterans with HSCR results. A majority of the sample was male (92.8%, n=5,356,442), White (76.2%, n=4,396,989), had served in the Army (53.0%, n=3,060,863), and had served in conflicts other than operations in Iraq or Afghanistan (88.8%, n=5,126,393); mean age was 61.1 years (SD=16.6; median = 64.0).

2.4. Evaluation

Descriptive statistics were computed for demographic and HSCR variables, including current living situation and referral acceptance in cases of positive screens. The criterion validity of the HSCR was evaluated by comparing the positive screens for current or risk of housing instability with the current living situation. Among veterans who screened positive for current or risk of housing instability, presence of administrative evidence for homelessness was compared between those who accepted or declined a referral for services using logistic regression analyses in order to evaluate screening and referral effectiveness in connecting veterans to services.

3. Results

3.1. Results of the HSCR

Results of initial screening showed that 0.8% (n=45,282) were positive for current housing instability, 1.0% (n=54,882) were positive for risk of housing instability, and

98.2% (n=5,671,332) screened negative (see Table 1). Among those who screened positive for housing instability, 61.9% were living in a homelessness situation, 24.5% were not; for those who screened positive for risk, 25.1% were living in a homeless situation, 65.3% were not. Results were similar whether we excluded or retained veterans without a current homeless living situation from the group that screened positive for housing instability. Administrative evidence for homelessness significantly varied depending on whether veterans accepted or declined a referral for services, with 61.3% of positive screens for housing instability who accepted services showing administrative evidence for homelessness as compared to only 19.0% who declined ($p<.01$); a similar difference was observed for other groups ($p<.01$).

Table 1. Homelessness Screening Clinical Reminder, Living Situation, Acceptance of Referral, and Administrative Evidence of Homelessness within 90 Days of Screening.

	Group 1: Housing Instability N=45,282 (0.8%)			**Group 2: Risk of Housing Instability N=54,882 (1.0%)**	**Group 3: Negative N=5,671,332 (98.2%)**	**Total**
Living Situation						
Homeless	27,878 (61.9%)			13,768 (25.1%)		41,646 (41.5%)
Friend/Family	18,355 (40.8%)			12,720 (23.2%)		31,075 (31.1%)
Shelter	2,091 (4.6%)			177 (0.3%)		2,268 (2.3%)
Street	4,791 (11.0%)			307 (0.6%)		5,278 (5.3%)
Motel/Hotel	2,461 (5.5%)			564 (1.0%)		3,025 (3.0%)
Non-Homeless	11,106 (24.5%)			35,843 (65.3%)		46,949 (46.8%)
Subsidized Housing	971 (2.2%)			3,110 (5.7%)		4,081 (4.1%)
Unsub. Housing	9,179 (20.4%)			32,395 (59.0%)		41,574 (41.6%)
Institution	866 (1.9%)			338 (0.6%)		1,204 (1.2%)
Unknown	6,127 (13.6%)			5,269 (9.6%)		11,396 (11.4%)
	Group 1: Overall	**Group 1a: Homeless Living Situation N=27,878**	**Group 1b: Non-Homeless Living Situation N=11,106**			
Accepted Referral for Service	28,279 (65.6%)	18,073 (67.9%)	6,664 (62.4%)	31,868 (60.5%)		60,147 (62.8%)
Administrative Evidence of Homelessness	21,502 (47.5%)	14,444 (51.8%)	4,017 (36.5%)	12,129 (22.1%)	43,955 (0.8%)	77,586 (1.3%)
If Accepted Referral	17,336 (61.3%)	11,755 (65.0%)	3,381 (50.7%)	10,054 (31.5%)		27,390 (45.8%)
If Declined Referral	2,922 (19.0%)	1,994 (23.3%)	453 (11.6%)	1,462 (7.0%)		4,384 (12.3%)

4. Discussion

A majority (61.9%) of veterans who self-identified or screened positive for current housing instability were indeed living in a homeless situation (e.g., shelter, street, hotel/motel, or doubled-up with friends or family) and conversely, a majority of those who screened positive for risk of housing instability were not currently living in a homeless situation (65.3%) (e.g., subsidized or unsubsidized housing, an institution). These results speak to the criterion validity of the HSCR in terms of differentiating between homelessness and risk of homelessness. Additionally, as evidenced by receipt

of administrative homelessness codes within 90 days after screening, veterans who screened positive for current housing instability and accepted a referral for services were more likely to receive homelessness prevention and intervention services (61.3%) than those who declined the referral (19.0%). Similarly, administrative evidence for homelessness was apparent for 31.5% of those with a positive screen for risk of housing instability who accepted a referral for services, as compared to only 7.0% for those who declined the referral. Thus, screening for housing instability and risk, which results in an acceptance of a referral for services in cases of positive screens, in turn leads to provision of homelessness-related services at a higher level, as evidenced by administrative data. A limitation of this study is that living situation was self-reported and those who screen negative are not asked to report their living situation, thus full diagnostic statistics including sensitivity and specificity cannot be computed.

5. Conclusion

The results of this study of almost 6 million records from a healthcare system suggests that screening for current housing instability and risk can lead to earlier identification, which can then lead to earlier referral for service provision. If such referrals are accepted, service provision is more likely to occur, as evidenced by administrative documentation of homelessness prevention and intervention services.

Acknowledgement

Funding provided by VA Health Services Research & Development grants IIR-13-334 and IIR 12-084. The views expressed are those of the authors and do not necessarily reflect the position or policy of the VA or the United States Government.

References

[1] Burt MR, Pearson CL, Montgomery AE. Community-wide strategies for preventing homelessness: Recent evidence. *Journal of Primary Prevention.* 2007;28, 213–228.
[2] Culhane DP, Metraux S, Byrne T. A prevention-centered approach to homelessness assistance: A paradigm shift? *Housing Policy Debate.* 2011;21(2):295–315.
[3] Montgomery AE, Fargo JD, Byrne TH, Kane V, Culhane DP. Universal screening for homelessness and risk for homelessness in the Veterans Health Administration. *American Journal of Public Health.* 2013;103(S2), S210–S211. doi:10.2105/AJPH.2013.301398
[4] Montgomery AE, Fargo JD, Kane V, Culhane DP. Development and validation of an instrument to assess imminent risk of homelessness among Veterans. *Public Health Reports.* 2014;129, 439–447.
[5] Montgomery AE, Dichter ME, Thomasson AM, Roberts CB. Services receipt following Veteran outpatients' positive screen for homelessness. *American Journal of Preventive Medicine.* 2016;50(3), 336–343. doi:10.1016/j.amepre.2015.06.035
[6] U.S. Department of Housing and Urban Development. Homeless emergency assistance and rapid transition to housing (HEARTH): defining "homeless" final rule. 2012. Retrieved from https://www.hudexchange.info/resource/1928/hearth-defining-homeless-final-rule/
[7] US Department of Veterans Affairs, VA Informatics and Computing Infrastructure (VINCI), in, US Department of Veterans Affairs, Washington DC, 2016.
[8] Peterson R, Gundlapalli AV, Carter ME, Palmer M, Redd A, Samore MH, Metraux S, Fargo JD. Identifying homelessness among Veterans using VA administrative data: Opportunities to expand detection criteria. *Plos ONE.* 2015;10(7):1-14; e0132664.doi:10.1371 /journal.pone.0132664.

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-579

Evidence for Busines Intelligence in Health Care: A Literature Review

Liz LOEWEN[a,1] and Abdul ROUDSARI[a]
[a] *School of Health Information Science, University of Victoria, Canada*

Abstract. This paper outlines a systematic literature review undertaken to establish current evidence regarding the impact of Business Intelligence (BI) on health system decision making and organizational performance. The review also examined BI implementation factors contributing to these constructs. Following the systematic review, inductive content analysis was used to categorize themes within the eight articles identified. This study demonstrated there is little evidence based literature focused on BI impact on organizational decision making and performance within health care. There was evidence found that BI does improve decision making. Implementation success was found to be dependent on several factors, many of which relate to broader organizational culture and readiness.

Keywords. Business intelligence, health, decision making, analytics, outcomes

1. Introduction

Clinical information systems are improving healthcare delivery through legible documentation, improved information sharing, and alerts among other benefits[1]. In addition, access to real-time information from these systems can enable front line managers to make informed decisions to drive system improvements. This concept, referred to as Business Intelligence (BI) can be defined as *the use of specialized tools to collect, analyze, and present organizational data to operational leaders in user-friendly format(s) to support organizational objectives*. BI is an emerging focus within the health sector but has become an established management practice in other sectors such as business, manufacturing, and finance where BI is viewed as a key component of strategic and operational decision making[2,3].

In all sectors, being rich in organizational data does not correlate directly to good information and, despite best intentions, "the problem is that most companies are not succeeding in turning data into knowledge and then results"[4, p. 118]. While there are few empirical studies examining the impact of BI in the health sector, there are numerous articles identifying the anticipated benefits of BI reinforcing the need for study in this area[5]. Expected benefits include: easier access to data[6-8]; time savings[7]; improved decision making[6]; improved outcomes[7]; and improved financial performance[9].

[1] Corresponding Author: Liz Loewen, PhD Candidate, School of Health Information Science, University of Victoria, HSD Building A202, Victoria, BC, Canada; E-mail:lloewen@uvic.ca.

2. Review Objectives

This systematic literature review was undertaken to establish the evidence demonstrating the impact of BI on decision making and organizational performance and to identify success factors for BI implementation in the health care sector. The review focused on the following questions: 1) What evidence exists that use of BI improves nurse or other health system manager decision making in health care? 2)What evidence exists that use of BI improves organizational performance in health care? 3) What are implementation success factors for BI in health care?

3. Methodology

The PRISMA methodology was used to guide the selection of articles and to structure both the screening and qualitative/quantitative assessment of included papers[10,11]. Search terms included: health care *or* medicine; BI *or* business analytics *or* big data; decision making – manager *or* nurse manager; organizational performance *or* outcomes; and implementation success factors. Databases were selected in consultation with subject matter experts to identify those common to the health sector, business and informatics and searches were conducted in July 2015.

Inclusion criteria were intentionally broad and included: English text; publication year>=2000 to reflect mature underlying health information systems; evidence based or existing systematic review; and, health system management related. Exclusion criteria were: clinical decision support (tools for individual patient care decisions); general decision making that did not reference or consider the underlying information systems; secondary or retrospective analysis; and articles without a research basis. Of the latter, numerous articles indicated a case study methodology however these were excluded where formal methods or controls for bias were not described. In total, 10 databases and search engines were included (See Table 1).

Table 1. Databases and search engines.

Database	Number of Results Returned
CINAHL with full text EBSCO	391
Medline with full text EBSCO	962
PubMed	265
Business Source Complete EBSCO	70
Web of Science Core Collection	577
IEEE Xplore Digital Library	407
Science Direct	139
Health Technology Assessments EBSCO	14
ACM Digital Library	12
Total	**2,837**

3.1. Screening and Quality Assessment

Searches were imported into EndNote™ reference management software and databases were merged and duplicates (n=547) were removed resulting in 2,290 unique articles. Articles were screened by title and abstract by one author (LL) and the remaining 342 articles were reviewed in detail along with an additional 36 identified through hand searching. Reviews were then validated with the second author (AR) with final determination based on consensus between the reviewers.

3.2. Data Analysis

Included articles were reviewed for quality and underlying bias and then analyzed using qualitative inductive content analysis[12]. Findings were marked manually and then transcribed into a graphical format using brainstorming features of Visio. The quantitative results were descriptive and, for the most part, not directly attributed to the presence of BI thus, the analysis focused on general themes and did not differentiate between anticipated versus actual impacts.

4. Results

Following screening, eight articles met the inclusion criteria (see Figure 1). The articles were predominantly descriptive with subjects reporting anticipated versus actual benefits with BI systems.

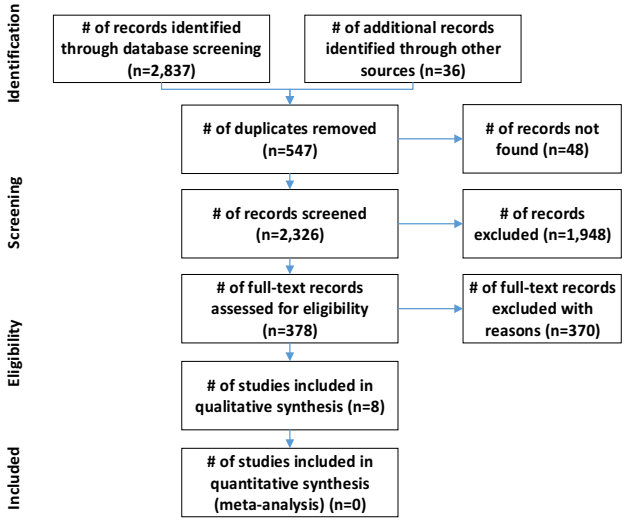


Figure 1. Search summary

Of the eight articles, seven directly examined health sector findings while one included health sector respondents among other service industries. Methods were predominantly qualitative and included: descriptive survey (n=1)[17], framework development methodology supported by case study (n=1)[18], mixed methods (n=2)[15,14], case study (n=3)[13,16,19], and systematic literature review (n=1)[20]. Four were published in 2014, and one in each of 2003, 2008, 2011 and 2015. Journals included: Communications of the Association for Information Systems; CIN: Computers, Informatics, Nursing; Journal of Ambulatory Care Management; Journal of Nursing Management; Journal of Oncology Practice; International Journal of Information Management (n=2); and International Journal of Accounting Information Systems.

Qualitative analysis of themes that presented in the articles resulted in six high level groupings: information needs/system indicators; information system quality; demonstrated/anticipated benefits; barriers to getting/using information; decision making impacts; and factors impacting BI adoption. While not all articles were

reflected in each of the theme groupings, each grouping contained concepts from at least four of the articles. Findings specific to each research question are summarized below (see Table 2).

Table 2. Summary of findings

Question	Findings
Evidence that use of BI improves manager decision making?	Manager reported improvements in decision making, economic awareness, ability to explain variances[14]
	Anticipated improvements if BI available: increased confidence in decisions, less subjective, more timely[13]
Evidence exists that use of BI improves organizational performance in health care?	Improved internal business processes (efficiency, customer intelligence) and ability to realize organizational objectives (enhanced profits, improved inventory turnover, partner relations) in all sectors although weaker in service sectors[15]
	Perception that organizational objectives such as length of stay were not being managed as effectively as they could be impacting quality of care and cost[13]
	Unit level improvements included reduced overtime and extra staffing hours (as compared with control units); managers reported better understanding of interrelated factors such as patient acuity, staffing and cost of care[14]
	Decreased morbidity and mortality, shorter wait times and length of stay and decreased cost[16]
Implementation success factors for BI in health care? *(often identified as gaps to address)*	Organizational: lack of skilled analytics resources[13,17,18]; leverage drivers such as external compliance or reporting mandates[17]; strong organizational vision[18]; address organizational silos[17]; and, address underlying care coordination factors[14,19]
	Technical: integration across multiple platforms[17,18]; and, need for a strong underlying technology platform[13,18]
	Data: underlying data quality and semantic interoperability systems[13,16,18]
	End user adoption: strong perceived usefulness and ease of use[14,19,20]; presentation of data that is meaningful and can be changed/controlled by end users[14,19,20]; and, ability to address fear of measurement and transparent reporting[14,19,20]

4.1. Discussion

The articles found through this search, while limited, do provide evidence for the research questions posed. The articles as a group suggest there is evidence that BI, when available to health system leaders and managers, would be used to inform decision making resulting in improved organizational performance. Given the prospective approach of several of the articles, it is worth noting the observation in Foshay and Kuziemsky[13] that there is a risk that managers may lack the skills to use the data for decision making even if it was available to them. The limited number of empirical articles found in this search on health sector BI is, in part, because the BI field itself is an emerging area[5].

4.2. Limitations

While the search terms were intentionally broad, it is still possible that articles were missed due to the emerging nature of work in this field and the potential that published research may be using alternative subject key words. The primary reviewer worked independently on the original reviews and may have missed key articles however, this is mitigated somewhat through the inclusion of new articles found through hand search and reference review.

5. Conclusion

These findings reinforce the need for research looking at BI impacts in the health care sector given the unique nature of health service delivery and its complex supporting organizational structures. The articles lend compelling arguments for the potential for BI to add value to health system manager practice. They also reinforce the need for research approaches that extend beyond implementation and user acceptance to look at organizational factors and the realization of organizational performance improvements.

References

[1] Han JE, Rabinovich M, Abraham P, Satyanarayana P, Liao TV, Udoji TN, Cotsonis GA, Honig EG, Martin GS. Effect of Electronic Health Record Implementation in Critical Care on Survival and Medication Errors. The American journal of the medical sciences 2016;**351**(6):576-581.
[2] Foshay N, Taylor A, Mukherjee A. Winning the hearts and minds of business intelligence users: the role of metadata. Information Systems Management 2014;**31**(2):167-180.
[3] Wixom B, Watson H. An Empirical Investigation of the Factors Affecting Data Warehousing Success. MIS Quarterly 2001;**25**(1):17-41.
[4] Davenport TH, Harris JG, De Long DW, Jacobson AL. Data to Knowledge to Results: Building an Analytic Capability. California management review 2001;**43**(2):117-138.
[5] Jourdan Z, Rainer RK, Marshall TE. Business intelligence: An analysis of the literature 1. Information Systems Management 2008;**25**(2):121-131.
[6] Bonney W. Applicability of Business Intelligence in Electronic Health Record. Procedia - Social and Behavioral Sciences 2013;**73**(0):257-262.
[7] Ferranti JM, Langman MK, Tanaka D, McCall J, Ahmad A. Bridging the gap: leveraging business intelligence tools in support of patient safety and financial effectiveness. Journal Of The American Medical Informatics Association: JAMIA 2010;**17**(2):136-143.
[8] Karami M, Fatehi M, Torabi M, Langarizadeh M, Rahimi A, Safdari R. Enhance hospital performance from intellectual capital to business intelligence. Radiology Management 2013;**35**(6):30-35.
[9] Glaser J, Stone J. Effective use of business intelligence. Healthcare Financial Management: Journal Of The Healthcare Financial Management Association 2008;**62**(2):68-72.
[10] Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. Annals of Internal Medicine 2009;**151**(4):264-269.
[11] Kable AK, Pich J, Maslin-Prothero SE. A structured approach to documenting a search strategy for publication: A 12 step guideline for authors. Nurse education today 2012;**32**(8):878-886.
[12] Seuring S, Gold S. Conducting content-analysis based literature reviews in supply chain management. Supply Chain Management: An International Journal 2012;**17**(5):544-555.
[13] Foshay N, Kuziemsky C. Towards an implementation framework for business intelligence in healthcare. International Journal of Information Management 2014;**34**(1):20-27.
[14] Ruland CM, Ravn IH. Usefulness and effects on costs and staff management of a nursing resource management information system. Journal of nursing management 2003;**11**(3):208-215.
[15] Elbashir MZ, Collier PA, Davern MJ. Measuring the effects of business intelligence systems: The relationship between business process and organizational performance. International Journal of Accounting Information Systems 2008;**9**(3):135-153.
[16] Ghosh B, Scott JE. Antecedents and Catalysts for Developing a Healthcare Analytic Capability. Communications of the Association for Information Systems 2011;**29**:395-410.
[17] Barkley R, Greenapple R, Whang J. Actionable data analytics in oncology: are we there yet? Journal Of Oncology Practice / American Society Of Clinical Oncology 2014;**10**(2):93-96.
[18] Brooks P, El-Gayar O, Sarnikar S. A framework for developing a domain specific business intelligence maturity model: Application to healthcare. International Journal of Information Management 2015;**35**(3):337-345.
[19] Ward CE, Morella L, Ashburner JM, Atlas SJ. An interactive, all-payer, multidomain primary care performance dashboard. The Journal Of Ambulatory Care Management 2014;**37**(4):339-348.
[20] Wilbanks BA, Langford PA. A review of dashboards for data analytics in nursing. Computers Informatics Nursing 2014;**32**(11):545-549.

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-584

Development of a Web-Based Quality Dashboard Including a Toolbox to Improve Pain Management in Dutch Intensive Care

Marie-José ROOS-BLOM[a,b,1], Wouter T. GUDE[a], Evert DE JONGE[b,c], Jan Jaap SPIJKSTRA[b,d], Sabine N. VAN DER VEER[e,f], Dave A. DONGELMANS[b,g], Nicolette F. DE KEIZER[a,b]

[a]*Dept. of Medical Informatics, Academic Medical Center, University of Amsterdam, The Netherlands*
[b]*National Intensive Care Evaluation (NICE) foundation, Amsterdam, The Netherlands*
[c]*Dept. of Intensive Care Medicine, Leiden University Medical Center, The Netherlands*
[d]*Dept. of Intensive Care Medicine, VU University Medical Center, Amsterdam, The Netherlands*
[e]*Health eResearch Centre, Farr Institute of Health Informatics Research, Manchester, UK*
[f]*Centre for Health Informatics, Division of Imaging, Informatics and Data Science, The University of Manchester, Manchester Academic Health Science Centre, UK*
[g]*Dept. of Intensive Care Medicine, Academic Medical Center, University of Amsterdam, The Netherlands*

Abstract. Audit and feedback (A&F) is a common strategy to improve quality of care. Meta-analyses have indicated that A&F may be more effective in realizing desired change when baseline performance is low, it is delivered by a supervisor or colleague, it is provided frequently and in a timely manner, it is delivered in both verbal and written formats, and it includes specific targets and an action plan. However, there is little information to guide operationalization of these factors. Researchers have consequently called for A&F interventions featuring well-described and carefully justified components, with their theoretical rationale made explicit. This paper describes the rationale and development of a quality dashboard including an improvement toolbox for four previous developed pain indicators, guided by Control Theory.

Keywords. Pain, Intensive Care, Quality Improvement, Toolbox, Control theory

1. Introduction

Audit and Feedback (A&F) is a common strategy to improve quality of care. It provides health professionals with a summary of their clinical performance over time [1] and is delivered to them in different formats such as a benchmark report or dashboard.

[1] Corresponding author, M. Roos-Blom, Dept. of Medical Informatics, University of Amsterdam, PO Box 22700, 1100 DE Amsterdam, The Netherlands; E-mail: m.blom@amc.nl

A Cochrane review showed that A&F is effective in improving provider compliance with desired practice, but with only a small median absolute improvement of 4.3% (interquartile range 0.5% to 16%) [2]. Meta-analyses have indicated that A&F may be more effective in realizing desired change when baseline performance is low, it is delivered by a supervisor or colleague, it is provided frequently and in a timely manner, it is delivered in both verbal and written formats, and it includes specific targets and an action plan [2-4]. However, there is little information to guide operationalization of these factors [5], which may lead to recommendations for improvement that are suboptimal [6]. Researchers have consequently called for A&F interventions featuring well-described and carefully justified components, with their theoretical rationale made explicit [1; 7].

Recently, our research group of the National Intensive Care Evaluation (NICE) quality registry [8] developed an A&F intervention that aims to improve clinical performance of Dutch intensive care units (ICUs) on pain management. Feedback is provided on four quality indicators through a web-based quality dashboard. We guided the design of our intervention by Control Theory and findings from a previous intervention undertaken by our research group among similar ICUs. A qualitative evaluation of the previous A&F intervention [9; 10] showed that ICUs experienced barriers to achieving improvement, including a lack of normative standards and benchmarks, inadequate case-mix adjustment, lack of knowledge on how to improve, and insufficient allocated time and staff [11]. To address all except the last barriers, we incorporated several strategies into the quality dashboard including explicit benchmark information, patient subgroup analyses, lists of patients who had not achieved the indicator target, and a quality improvement (QI) toolbox containing potential barriers and suggested actions to improve practice.

This paper describes the rationale and development of the QI toolbox for four previous developed pain indicators [12]. The effectiveness of the QI toolbox will be evaluated in a randomized controlled trial.

2. Methods

2.1. Theoretical Rationale

Our theoretical rationale of the mechanisms through which health professionals aim to improve their clinical performance is based on Control Theory as shown in Figure 1. Control Theory hypothesizes that behavior is goal-driven, and professionals are prompted to change behavior (e.g., execute improvement actions) when observing a negative discrepancy between their current performance and a goal, until the discrepancy is eliminated. Feedback reports are input for comparing perceived clinical performance to performance goals (see upper-left grey box in Figure 1) However, if professionals lack skills, knowledge or strategies for action, they may disengage from goal attainment and stop trying to achieve the goal [13]. Feedback may, therefore, be more effective when accompanied by action plans to inform behavioral adjustment to reduce discrepancy (see bottom-right grey box in Figure 1) [6]. Our QI toolbox will supply in action planning by supporting ICU professionals to translate their intentions into actions, and enhances the likelihood that actions will be completed by providing supporting material to execute those planned actions.

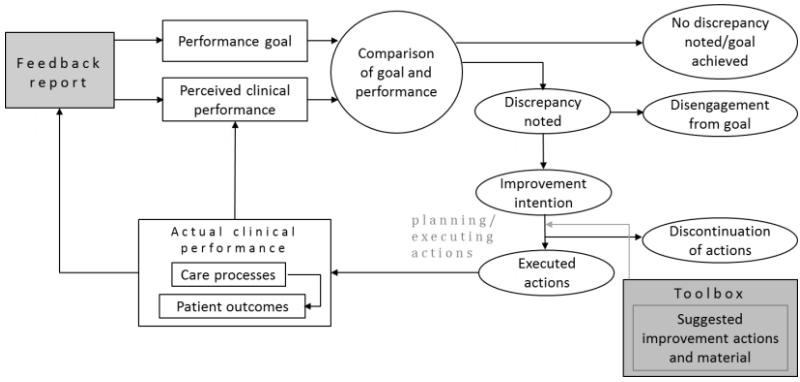


Figure 1. Illustration of hypothesized role played by feedback and the QI toolbox to promote development of improvement intentions and behaviour change. Adapted from Carver & Scheier's Control Theory.

2.2. Toolbox Development: Identification of Barriers and Improvement Actions

The quality dashboard provides feedback on the four previously developed quality indicators reflecting best practice for pain management: 1) perform pain measurements each shift; 2) achieve acceptable pain scores; 3) repeat pain measurements in case of unacceptable scores within one hour; 4) normalize unacceptable pain scores within one hour. To develop the QI toolbox we analyzed the pain management process in the ICU. The pain management process consists of three steps: 1) pain assessment, 2) medication or other treatment prescription, and 3) treatment effect evaluation. The Systems Engineering Initiative for Patient Safety (SEIPS) model provides a framework for understanding the structures, processes and outcomes in health care [14]. Based on the SEIPS model, two of the authors (MRB and JJS) identified potential barriers that could lead to poor performance on each of the four pain management quality indicators. Next, for each barrier we listed expert and guideline-based examples of goal-oriented actions that could be effective in attaining better performance. During an expert focus group the preliminary list of barriers and actions was discussed, prioritized and supplemented with other expert and practice-based actions. The expert panel consisted of two intensive care nurses; one hospital pharmacist; two anesthesiologist-intensivists; two internist-intensivists; and one research coordinator pain management. For the final list of actions we searched for supporting material in the literature. Additionally, we asked the expert panel for locally developed supporting material that could facilitate action implementation. These improvement actions and materials together with the quality indicators, were included in a web-based dashboard that will be offered to the ICUs. Potential end users critically assessed the dashboard for usability.

3. Results

The QI toolbox (Figure 2) consists of 18 unique barriers and 26 unique improvement actions. On average, the toolbox lists 12 barriers per indicator (range, 11 to 13) and 5 actions per barrier (range, 1 to 8). Six actions have supporting materials attached to facilitate its implementation, including posters, educational PowerPoint presentations, pocket cards, information leaflets and protocols. The toolbox is incorporated within the

action plan in the dashboard, see Figure 2. ICU teams can develop action plans for each quality indicator included in the dashboard. To do so, they can select or deselect potential barriers in their local care process organization, and pick actions. The toolbox displays actions that are associated with the selected barriers, and hides those that are only associated with deselected barriers. ICU teams can pick those actions they intend to undertake. Teams can also define their own barriers and actions. Each improvement action will be assigned by a deadline and responsible person(s).

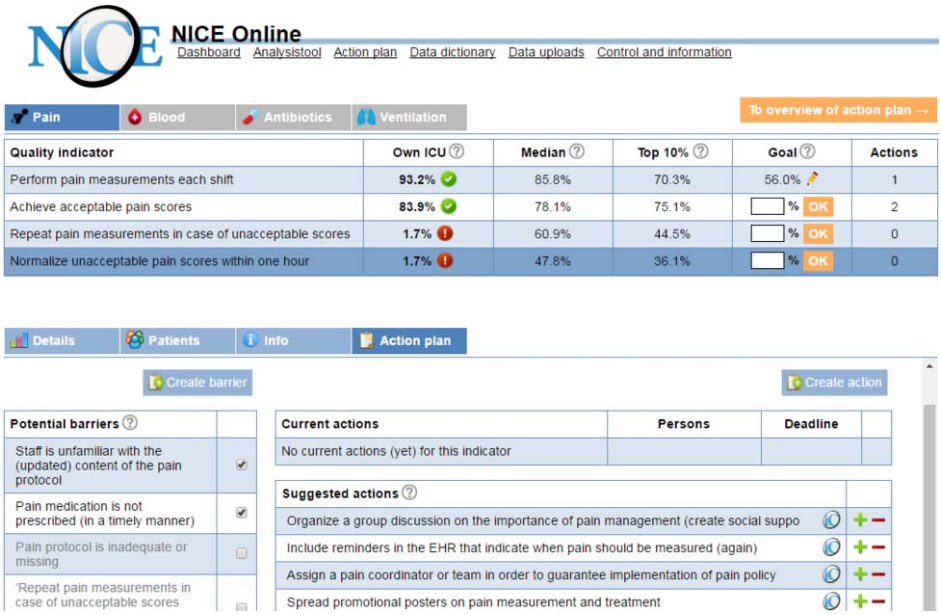


Figure 2. The NICE dashboard (translated from Dutch).

4. Discussion

We developed a web-based quality dashboard with QI toolbox, guided by Control Theory, to assist Dutch ICUs in achieving better performance on pain management.

Behavior change is most likely if feedback is accompanied by specific goals and action plans [6]. A previous trial of Ivers et al. [15] studying goal-setting and action planning did not lead to quality improvement. Their intervention lacked the inclusion of active, practice-based support and was therefore believed to be ineffective. Our QI toolbox does include expert-based actions complemented with evidence from the literature, resulting in a practice-based list of goal-oriented actions and supporting material which enhances its employment. Following Control Theory [13], we believe that our toolbox may reduce the intention-action gap by suggesting improvement actions in case they do not know what they can do to improve (lack of knowledge) and materials in case they do not know how to do it (lack of skills).

Our QI toolbox was designed to offer a comprehensive set of actions that are specific enough to be actionable [16], yet generic enough to be relevant for all ICUs. To overcome the problem of being confronted with a high number of inapplicable actions, e.g. because actions might have already been implemented or might not contribute to solving a local ICU's specific barriers, teams can select those barriers from the toolbox that are relevant to their own context; the toolbox displays only those

suggested actions associated with the selected barriers. Nevertheless, actions can still be too generic to guarantee swift implementation (e.g. increase efficiency of work process). We aim to continuously expand and improve the toolbox with new or revised actions and supporting materials to address this issue in the future.

We will evaluate the effectiveness of the QI toolbox in an upcoming head-to-head randomized controlled trial (ClinicalTrials.gov NCT02922101).

5. Conclusion

Our web-based quality dashboard is one of the first A&F interventions with a QI toolbox based on explicit theories. The toolbox offers a solution to ICU professionals in planning and executing more effective improvement strategies for pain management. Its effectiveness will be evaluated and if successful it will be applied to other areas of ICU care.

References

[1] N.M. Ivers, A. Sales, et al., No more 'business as usual' with audit and feedback interventions: towards an agenda for a reinvigorated intervention, *Implement Sci* **9** (2014), 14.
[2] N.M. Ivers, G. Jamtvedt, et al., Audit and feedback: effects on professional practice and healthcare outcomes, *Cochrane Database Syst Rev* **6** (2012), CD000259.
[3] N.M. Ivers, J.M. Grimshaw, et al., Growing literature, stagnant science? Systematic review, meta-regression and cumulative analysis of audit and feedback interventions in health care, *J Gen Intern Med* **29** (2014), 1534-1541.
[4] T. Sinuff, J. Muscedere, et al., A qualitative study of the variable effects of audit and feedback in the ICU, *BMJ Qual Saf* **24** (2015), 393-399.
[5] R. Foy, M.P. Eccles, et al., What do we know about how to do audit and feedback? Pitfalls in applying evidence from a systematic review, *BMC Health Serv Res* **5** (2005), 50.
[6] B. Gardner, C. Whittington, et al., Using theory to synthesise evidence from behaviour change interventions: the example of audit and feedback, *Soc Sci Med* **70** (2010), 1618-1625.
[7] H.L. Colquhoun, J.C. Brehaut, et al., A systematic review of the use of theory in randomized controlled trials of audit and feedback, *Implement Sci* **8** (2013), 66.
[8] N. van de Klundert, R. Holman, et al., Data Resource Profile: the Dutch National Intensive Care Evaluation (NICE) Registry of Admissions to Adult Intensive Care Units, *Int J Epidemiol* **44** (2015), 1850-1850h.
[9] M.L. de Vos, S.N. van der Veer, et al., A multifaceted feedback strategy alone does not improve the adherence to organizational guideline-based standards: a cluster randomized trial in intensive care, *Implement Sci* **10** (2015), 95.
[10] S.N. van der Veer, M.L. de Vos, et al., Effect of a Multifaceted Performance Feedback Strategy on Length of Stay Compared With Benchmark Reports Alone: A Cluster Randomized Trial in Intensive Care, *Crit Care Med* (2013).
[11] M.L. de Vos, S.N. van der Veer, et al., Process evaluation of a tailored multifaceted feedback program to improve the quality of intensive care by using quality indicators, *BMJ Qual Saf* **22** (2013), 233-241.
[12] M.J. Roos-Blom, W.T. Gude, et al., Development of actionable quality indicators and a quality improvement toolbox for pain management in Dutch ICUs., *Submitted* (2016).
[13] C.S. Carver and M.F. Scheier, Control theory: a useful conceptual framework for personality-social, clinical, and health psychology, *Psychol Bull* **92** (1982), 111-135.
[14] P. Carayon, A.S. Hundt, et al., Work system design for patient safety: the SEIPS model, *Quality & Safety in Health Care* **15** (2006), I50-I58.
[15] N.M. Ivers, K. Tu, et al., Feedback GAP: pragmatic, cluster-randomized trial of goal setting and action plans to increase the effectiveness of audit and feedback interventions in primary care, *Implement Sci* **8** (2013), 142.
[16] E.A. Locke and G.P. Latham, Building a practically useful theory of goal setting and task motivation. A 35-year odyssey, *Am Psychol* **57** (2002), 705-717.

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-589

A Digital Framework to Support Providers and Patients in Diabetes Related Behavior Modification

Samina ABIDI[a,1], Michael VALLIS[b], Helena PICCININI-VALLIS[c], Syed Ali IMRAN[d], Syed Sibte Raza ABIDI[e]
[a] *Medical Informatics, Faculty of Medicine, Dalhouise University, Canada*
[b] *NSHA Behavior Change Institute, QEII Health Sciences Center, Halifax, Canada*
[c] *Department of Family Medicine, Faculty of Medicine, Dalhousie University, Canada*
[d] *Division of Endocrinology and Metabolism, Dalhousie University, Canada*
[e] *NICHE Research Group, Faculty of Computer Science, Dalhousie University, Canada*

Abstract. We present Diabetes Web-Centric Information and Support Environment (D-WISE) that features: (a) Decision support tool to assist family physicians to administer Behavior Modification (BM) strategies to patients; and (b) Patient BM application that offers BM strategies and motivational interventions to engage patients. We take a knowledge management approach, using semantic web technologies, to model the social cognition theory constructs, Canadian diabetes guidelines and BM protocols used locally, in terms of a BM ontology that drives the BM decision support to physicians and BM strategy adherence monitoring and messaging to patients. We present the qualitative analysis of D-WISE usability by both physicians and patients

Keywords. Behaviour Modification, Self-Management, Diabetes, Knowledge Management, Decision Support, Ontology, Semantic Web

1. Introduction

The Canadian diabetes clinical practice guideline specifically recommends that diabetes patients should be assisted to self-manage their disease with the help of an integrated diabetes team. Diabetes in Canada is managed by Family Physicians (FP) and certified diabetes educators. At the provider level, the challenge is that they are not trained to administer behaviour modification strategies to patients. At the patient level, the challenge is the low uptake of the educational material provided to patients—some patients do not comply with the recommendations and others do not have the self-efficacy or motivation to pursue the recommendations. We believe that a informatics based behaviour modification environment can (a) empower providers to administer behaviour modification by helping them design personalized behaviour modification strategy based on the patient's specific needs; and (b) motivate patients to adhere to their behaviour modification strategy by monitoring and messaging so that they adopt healthy behaviours and to achieve efficacy to self-manage their condition at home.

[1] Corresponding Author: Samina Abidi, Department of Community Health and Epidemiology, Centre for Clinical Research, 5790 University Avenue, Halifax, NS B3H 1V7, Canada. Email: samina.abidi@dal.ca

In this paper, we present "Diabetes Web-Centric Information and Support Environment" (D-WISE) that features the following functionalities: (i) Assessment of FP readiness to administer behavior change interventions to patients; (ii) Behavior modification (BM) educational support to FP so that they can offer behavior modification (BM) interventions to patients; (iii) Personalized self-management programs and education to help patients modify behaviors; (iv) Monitoring the patient's progress as per their behavior change program and motivating them to comply with it. D-WISE offers these functionalities through an interactive web-based interface to physicians, whereas the patient's self-management program and associated behavior interventions are delivered through smart mobile devices [1]. D-WISE leverages semantic web technologies, where it incorporates an BM ontology that represents evidence-based BM theories that can be reasoned over using a patient profile to generate a personalized BM strategy [2].

2. Solution Approach

The theoretical foundation of D-WISE is grounded in Behavior Change Models (the knowledge content) and Healthcare Knowledge Management (the knowledge translation method). We have modeled and computerized the constructs of Social Cognitive Theory (SCT) [3] and the BM protocols used by the Halifax Behaviour Change Institute (BCI) in terms of a BM ontology that is used to develop behavior profiles for both FP and patients. Using the knowledge represented within the BM ontology our behavior modification approach is to: (i) assess FP's readiness to administer behavior modification counselling to patients; (ii) guide FP to assess patient's readiness and self-efficacy and then design a personalized behavior modification interventions in a shared-decision making setting—patients set short-term behaviour goals and design a feasible action plan; (iii) motivate patients to achieve their goals through motivational messaging and showing them their progress.

3. D-WISE Implementation Methodology

3.1. Development of Behavior Modification (BM) Algorithm

The BM protocols applied at the BCI were translated into high-level BM algorithms that captured the readiness assessment tools, range of provider and patients' inputs, sequence of the BM strategy and corresponding educational material. The BM algorithm compute: (i) BM Readiness Assessment in terms of Ready, Ambivalent and Not ready; (ii) BM Decisional balance scored ranging from 1 (completely disagree) to 10 (completely agree); (iii) BM Self-efficacy assessment of individuals.

3.2. Modeling of Behavior Modification Knowledge

The BM Ontology was developed using the Web Ontology Language (OWL), applying ontology modularization principles, with two main ontological modules:

Information Personalization Module creates unique patient profiles that consist of four elements: (i) *Medical Profile* represents patient's current medical data; (ii)

Readiness Assessment represents the readiness assessment of FP to administer BM to patients, and for patients to uptake BM interventions; (iii) *Decisional Balance Assessment* measures positive and negative perceptions of FP and patients who are not ready or are ambivalent towards BM; and (iv) *Self-efficacy Assessment* assesses SCT based self-efficacy of the FP and the patient in providing/adhering to BM interventions. Readiness, decisional balance and self-efficacy assessments required the ontological modeling of respective paper-based assessment questionnaires. Information Personalization Module contains 9 classes (Barrier, Barrier to Change, Clinical Profile, Decisional Balance Result, Readiness Assessment, Readiness Assessment Result, Self Efficacy Questionnaire, Self Efficacy Result, Prognostic Factors).

Domain Knowledge Module represents disease-specific knowledge and self-management knowledge, e.g. barriers to diabetes self-management and BM support materials, and goal setting support. Domain Knowledge Module contains 8 classes (Clinical Parameter, CPG Recommendation, Behavior Recommendation, Feedback After Assessment, Goal Setting, Reminder Schedule, Target Behavior and Strategies). BMO represents 18 top-level classes, 16 top-level object properties. The knowledge was validated for clinical correctness by the domain experts.

3.3. D-WISE Framework

D-WISE is a web-based decision support framework with specialized tools supporting the FP and diabetes patient to achieve BM in a standardized and sustained manner [1]. D-WISE FP tool firstly provides readiness assessment questionaries' to assess the readiness and self-efficacy of the FP to administer BM. The readiness assessment logic captured within the BM ontology is leveraged to dynamically select and adapt the standard assessment tools suit the context of the FP. For FP who are assessed to be 'ready', D-WISE secondly establishes a shared decision making session between the patient and the FP to help them design a patient-specific BM strategy; this functionality is akin to the BM consultations performed at the BCI.

D-WISE patient tool, implemented as both a web-based system and an Android mobile app, supports the patient to achieve BM and diabetes self-management through interfaces that allow goal setting, behavior shaping, stimulus control and reinforcement management based on the SCT constructs modeled by the BM ontology. These BM elements are monitored via a patient diary that allows patients to report vitals, diet, exercise, stress and mood. The patient-reported data is analyzed to establish their progress with respect to their BM strategy, and to keep them engaged motivational messages (and alerts) are periodically sent to encourage patients to adhere to their selected BM goal and the BM strategy.

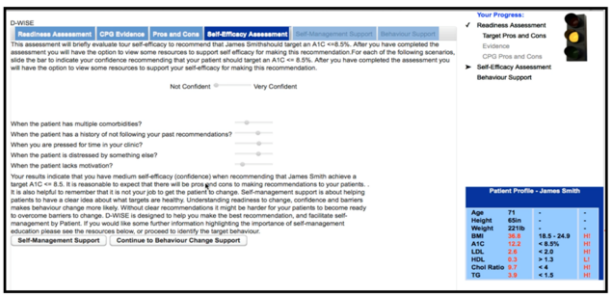


Figure 1: Provider tool assessing self-efficacy of the FP/CDE

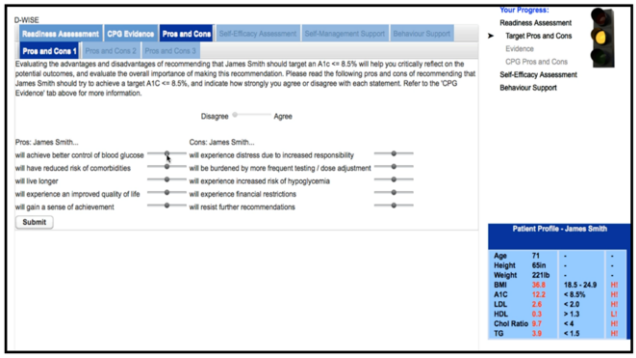


Figure **2**: Provider tool assessing BM pros and cons

4. Evaluation of D-WISE

To measure whether D-WISE meets the functional goals, content suitability and usability needs of FP and patients, we conducted two pilot studies that employed the Think Aloud Protocol (TAP) [4] using a cognitive and usability engineering framework [4]. We recruited 10 licensed FP for the provider study and 11 adult patients for patient study following ethics approval. The sample size estimate is based on the evidence that 70% of severe usability problems can be uncovered within the first five users, and up to 85% by the eighth user [5]. Each FP was provided with 3 standard case scenarios that simulated 3 different patients thus yielding 30 TAP. Patients were presented with a standard behavioral recommendation (such as increase physical activity) that yielded 11 TAP. The study participants interacted with D-WISE and were encouraged to think aloud during their interactions; their screen activity and audio was recorded and analyzed in the ATLAS.Ti software using the inductive thematic coding method.

4.1. Qualitative Results from Provider Tool Study

In total 22 independent open codes based on usability issues were identified in 30 provider TAP (see figure 3). A number of these codes were highly grounded in the data e.g. “Need more patient information for pros and cons” is grounded in 19 quotations and “Need more information for readiness assessment” is grounded in 11 quotations. In total 17 themes of axial categories of usability problems emerged. In conclusion, most problems were associated with navigation, and some to the presentation and understanding of the content. The FP were satisfied with the BM content of D-WISE.

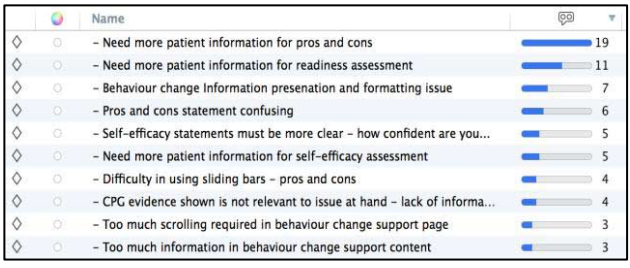


Figure **3**: Open codes observed in Provider TAP

4.2. Qualitative Results from Patient Tool Study

Patient TAP yielded 17 open codes and 9 axial categories (see figure 4). Out of 17 codes, most critical codes include: 'Unsure of goal setting data entry field' which was grounded in 11 quotations, 'Sliding bar problems' grounded in 7 quotations and 'Problems with scrolling' grounded in 6 quotations. It was concluded that the patients were satisfied with the BM content and purpose of D-WISE.

Figure **4**: Open and axial codes derived from patient TAP

5. Discussion and Concluding Remarks

We presented a digital health approach that translates BM models to a point-of-care BM tool. The key contribution of our approach is the modeling of BM knowledge and then operationalization of the BM model to generate personalized BM strategies for diabetes patients to self-manage their condition. Furthermore, D-WISE presents a unique BM environment that provides a shared decision making opportunity to both providers and patients to administer personalized BM interventions based on theoretical BM constructs. D-WISE assesses the readiness of patients and then accordingly guides patients to specify their barriers and goals, thus ensuring that adherence to the BM strategy is feasible. The BM knowledge model and strategy formulation approach is scalable in nature, such that can be readily applied other chronic diseases. D-WISE will now be evaluated for its efficacy to influence BM.

References

[1] S. Abidi, M. Vallis, SSR ABIDI, H. Piccinini-Vallis, SA. Imran. D-WISE: Diabetes web-centric information and support environment: Conceptual specification and proposed evaluation. Canadian Journal of Diabetes, 2014, 38(3), pp. 205-11. Elsevier

[2] SSR. ABIDI, S. Abidi. An ontology-driven personalization framework for designing theory-driven self-management interventions. Process Support and Knowledge Representation in Health Care., 2013, pp. 97-112, Springer International Publishing.

[3] A. Bandura. (1986) Social foundations of thought and action; A social cognitive theory. Prentice Hall: Englewood Cliffa, N.J.

[4] AW. Kushnirk, VL. Patel. Cognitive evaluation of decision making processes and assessment of information technology in medicine. International Journal of Medical Informatics, (1998);51(2-3);83-90

[5] AW. Kushnirk, VL Patel. Cognitive and usability engineering methods for the evaluation of clinical information systems. J Biomed Informatics, (2004);37(1):56-76

This page intentionally left blank

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

Subject Index

This page intentionally left blank

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

Author Index

This page intentionally left blank