Health Informatics and Knowledge Management 2014

Proceedings of the Seventh Australasian Workshop on Health Informatics and Knowledge Management (HIKM 2014), Auckland, New Zealand, 20 - 23 January 2014

Jim Warren and Kathleen Gray, Eds.

Volume 153 in the Conferences in Research and Practice in Information Technology Series. Published by the Australian Computer Society Inc.

Published in association with the ACM Digital Library.
Table of Contents

Proceedings of the Seventh Australasian Workshop on Health Informatics and Knowledge Management (HIKM 2014), Auckland, New Zealand, 20 - 23 January 2014

Preface ............................................................... vii
Programme Committee ........................................... viii
Organising Committee ............................................ ix
Welcome from the Organising Committee ....................... x
CORE - Computing Research & Education ....................... xi
ACSW Conferences and the Australian Computer Science Communications ........................................... xiii
ACSW and HIKM 2014 Sponsors ................................... xv

Contributed Papers

Crowdsourcing for Clinical Research An Evaluation of Maturity .................................................. 3
  Kate E. Birch, Kayla J. Heffernan

Investigating the Potential of Interactive Media to Encourage Engagement with Type 1 Diabetes Management ................................................................. 13
  Alexander S. Borsboom

Social Media and Online Survey: Tools for Knowledge Management in Health Research ............... 21
  Mark Merolli, Fernando Jose Martin Sanchez, Kathleen Gray

BiomRKRS: A Biomarker Retrieval and Knowledge Reasoning System ........................................... 31
  Bahadorreza Ofoghi, Guillermo Hugo Lopez Campos, Karin Verspoor, Fernando Jose Martin Sanchez

Automatic Population of Structured Reports from Narrative Pathology Reports .......................... 41
  Ying Ou, Jon Patrick

What Pacific People Think of Online Mental Health Information .................................................... 51
  Bernadette Peni, Karen Day, Martin Orr

Health Intelligence: Discovering the Process Model Using Process Mining by Constructing Start-to-End Patient Journeys .................................................. 59
  Lua Perimal-Lewis, Denise de Vries, Campell H. Thompson

Developing High Risk Clusters for Chronic Disease Events with Classification Association Rule Mining 69
  Shen Song, Jim Warren, Patricia Riddle

Building an Empirical Treatment Protocol from High-Resolution Traumatic Brain Injury Data .......... 79
  Anthony Stell, Laura Moss, Ian Piper

Prescription Patterns for Pacific Patients with Poor Cardiovascular Disease Medication Supply ........ 89
  Jim Warren, Yulong Gu, John Kennelly, Tai Boyce

Author Index ........................................................... 97
Preface

We are pleased to present the papers from the Seventh Australasian Workshop on Health Informatics and Knowledge Management (HIKM 2014), held on 22–23 January at Auckland University of Technology (AUT) in Auckland as part of Australasian Computer Science Week.

HIKM provides a venue for Computer Science researchers who would not necessarily identify with the ‘Health Informatics’ community to present applications of their technical expertise to healthcare, while simultaneously acting as a forum for confirmed Health Informatics experts to share their latest contributions with a tech-savvy audience. It is always the hope that this will encourage a few more computer scientists to develop a deeper interest in the IT challenges of the health sector.

While ‘Big Data’ was not an advertised theme of the workshop, it’s a little hard to ignore the amount of hype that is building around this term. Those of us that have been analysing health data for a little while have long recognised its capacity to inform us about trends, and deficiencies, in healthcare delivery, to inform decision support at the levels of public health and local quality improvement, as well as to inform clinical decision support for the point of care. Conversely, we are well aware of the challenges of working with healthcare data, such as those related to sparseness or non-uniformity of clinical coding. Nonetheless, there is some validity in the Big Data excitement as the data sources are rapidly broadening out from ‘traditional’ repositories such as laboratory tests and dispensing records, to encompass home telemonitoring and ambient sources such as GPS and social media content, not to mention genomic data. But other broader trends also drive our work, not the least of which is the non-sustainability of present healthcare delivery in light of growing burden of chronic illness, ageing population, limited economic growth and the difficulty of fielding a sufficient health workforce – these challenges create a ‘burning platform’ with IT-enabled innovation in healthcare delivery as a key strategy for avoiding the unsatisfactory conclusions of maintaining ‘business as usual.’

This year HIKM received 21 submissions. After peer review of the full manuscripts, 10 submissions were accepted for presentation at HIKM and inclusion in the proceedings (a 48% acceptance rate). These papers provide insight on a range of promising avenues for improving health delivery through IT. In these papers we find methods as diverse as natural language processing, clustering and social media, and the pursuit of applications as diverse as improving treatment protocols and patient pathways, inferring structured data from free-text notes and helping health consumers to help themselves with online learning. They are published here as Volume 153 of Conferences in Research and Practice in Information Technology (CRPIT).

Jim Warren
University of Auckland

Kathleen Gray
University of Melbourne

HIKM 2014 Programme Chairs
January 2014
Programme Committee

Chairs

Jim Warren, University of Auckland
Kathleen Gray, University of Melbourne, Australia

Members

David Hansen, CSIRO, Australia
Anthony Maeder, University of Western Sydney, Australia
Farah Magrabi, University of New South Wales, Australia
Dave Parry, Auckland University of Technology, New Zealand
Tony Sahama, Queensland University of Technology, Australia
Jan Stanek, University of South Australia, Australia
Trish Williams, Edith Cowan University, Australia
Ping Yu, University of Wollongong, Australia

Additional Reviewers

The Programme Committee extends its thanks to the experts who provided their additional support to the peer review process by providing timely and insightful reviews on submitted manuscripts.

Dr Jim Basilakis, University of Western Sydney, Australia
Prof Gleb Beliakov, Deakin University, Australia
Dr Joanne Curry, University of Western Sydney, Australia
Dr Karen Day, University of Auckland, New Zealand
Dr Adam Dunn, University of New South Wales, Australia
Dr Randike Gajanayake, Queensland University of Technology, Australia
Dr Filippo Galgani, University of New South Wales, Australia
Daniel Grunwell, Queensland University of Technology, Australia
Dr Yulong Gu, University of Auckland, New Zealand
Dr Liwan Liyanage, University of Western Sydney, Australia
Mrs Elica Mehr, University of Auckland, New Zealand
Dr Mei-Sing Ong, University of New South Wales, Australia
Dr Chris Paton, Oxford University, United Kingdom
Dr Oscar Perez Concha, University of New South Wales, Australia
Ms Nouran Ragaban, University of Auckland, New Zealand
Dr Klaus Veil, University of Western Sydney, Australia
Dr Ying Wang, University of New South Wales, Australia
Organising Committee

Chairs
Tony Clear and Russel Pears

Venue
Tony Clear

Communications
Russel Pears, Hui Ling Tan, Melanie Curry-Irons and Ryan Butler

Finance
Alison Clear and Eva Ihaia

Sponsorship
Stephen Thorpe

Operations
Adam Winship and Eva Ihaia

Programme, proceedings and booklet
Alison Clear

Catering and registration
AUT Hospitality Services
Welcome from the Organising Committee

On behalf of the Organising Committee, it is our pleasure to welcome you to Auckland and to the 2014 Australasian Computer Science Week (ACSW 2014). Auckland is New Zealand’s largest urban area with a population of nearly one and a half million people. As the centre of commerce and industry, Auckland is the most vibrant, bustling and multicultural city in New Zealand. With the largest Polynesian population in the world, this cultural influence is reflected in many different aspects of city life. ACSW 2014 will be hosted at the City Campus of Auckland University of Technology (AUT), which is situated just up from the Town Hall and the Auckland central business district. ACSW is the premier event for Computer Science researchers in Australasia. ACSW2014 consists of conferences covering a wide range of topics in Computer Science and related areas, including:

- Australasian Computer Science Conference (ACSC) (Chaired by Bruce Thomas and Dave Parry)
- Australasian Computing Education Conference (ACE) (Chaired by Jacqueline Whalley and Daryl D’Souza)
- Australasian Information Security Conference (AISC) (Chaired by Udaya Parampalli and Ian Welch)
- Australasian User Interface Conference (AUIC) (Chaired by Burkhard C. Wünsche and Stefan Marks)
- Australasian Symposium on Parallel and Distributed Computing (AusPDC) (Chaired by Bahman Javadi and Saurabh Kumar Garg)
- Australasian Workshop on Health Informatics and Knowledge Management (HIKM) (Chaired by James Warren)
- Asia-Pacific Conference on Conceptual Modelling (APCCM) (Chaired by Georg Grossmann and Motoshi Saeki)
- Australasian Web Conference (AWC) (Chaired by Andrew Trotman)

This year reflects an increased emphasis for ACSW on community building. Complementing these published technical volumes therefore, ACSW also hosts two doctoral consortia and a number of associated workshops, including those for the Heads and Professors of Computer Science, plus for the first time the ‘Australasian Women in Computing Celebration’. Naturally in addition to the technical program, there are a range of events, which aim to provide the opportunity for interactions among our participants. A welcome reception will be held in the atrium of the award winning newly built Sir Paul Reeves Building, which has integrated the city campus as a hub for student activity and provides a wonderful showcase for this year’s ACSW. The conference banquet will be held on campus in one of the reception rooms in this impressive complex.

Organising a multi-conference event such as ACSW is a challenging process even with many hands helping to distribute the workload, and actively cooperating to bring the events to fruition. This year has been no exception. We would like to share with you our gratitude towards all members of the organising committee for their combined efforts and dedication to the success of ACSW2014. We also thank all conference co-chairs and reviewers, for putting together the conference programs which are the heart of ACSW, and to the organisers of the symposia, workshops, poster sessions and accompanying conferences. Special thanks to Alex Potanin, as the steering committee chair who shared valuable experiences in organising ACSW and to John Grundy as chair of CoRE for his support for the innovations we have introduced this year. We’d also like to thank Hospitality Services from AUT, for their dedication and their efforts in conference registration, venue, catering and event organisation. This year we have secured generous support from several sponsors to help defray the costs of the event and we thank them for their welcome contributions.

Last, but not least, we would like to thank all speakers, participants and attendees, and we look forward to several days of stimulating presentations, debates, friendly interactions and thoughtful discussions.

We hope your stay here will be both rewarding and memorable, and encourage you to take the time while in New Zealand to see some more of our beautiful country.

Tony Clear
Russel Pears
School of Computer & Mathematical Sciences
ACSW2014 General Co-Chairs
January, 2014
CORE welcomes all delegates to ACSW2014 in Auckland. CORE, the peak body representing academic computer science in Australia and New Zealand, is responsible for the annual ACSW series of meetings, which are a unique opportunity for our community to network and to discuss research and topics of mutual interest. The component conferences of ACSW have changed over time with additions and subtractions ACSC, ACE, AISC, AUIC, AusPDC, HIKM, ACDC, APCCM, CATS and AWC have now been joined by the Australasian women in computing celebration (AWIC), two doctoral consortia (ACDC and ACE-DC) and an Australasian Early Career Researchers Workshop (AECRW) which reflect the evolving dimensions of ACSW and build on the diversity of the Australasian computing community.

In 2014, we have again chosen to feature a small number of keynote speakers from across the discipline: Anthony Robins (ACE), John Mylopoulos (APCCM), and Peter Gutmann (AISC). I thank them for their contributions to ACSW2014. The efforts of the conference chairs and their program committees have led to strong programs in all the conferences, thanks very much for all your efforts. Thanks are particularly due to Tony Clear, Russel Pears and their colleagues for organising what promises to be a vibrant event. Below I outline some of CORE's activities in 2012/13.

I welcome feedback on these including other activities you think CORE should be active in.

The major sponsor of Australian Computer Science Week:
- The venue for the annual Heads and Professors meeting
- An opportunity for Australian & NZ computing staff and postgrads to network and help develop their research and teaching
- Substantial discounts for attendees from member departments
- A doctoral consortium at which postgrads can seek external expertise for their research
- An Early Career Research forum to provide ECRs input into their development

Sponsor of several research, teaching and service awards:
- Chris Wallace award for Distinguished Research Contribution
- CORE Teaching Award
- Australasian Distinguished Doctoral Dissertation
- John Hughes Distinguished Service Award
- Various Best Student Paper awards at ACSW

Development, maintenance, and publication of the CORE conference and journal rankings. In 2013 this includes a new portal with a range of holistic venue information and a community update of the CORE 2009 conference rankings.

Input into a number of community resources and issues of interest:
- Development of an agreed national curriculum defining Computer Science, Software Engineering, and Information Technology
- A central point for discussion of community issues such as research standards
- Various submissions on behalf of Computer Science Departments and Academics to relevant government and industry bodies, including recently on Australian Workplace ICT Skills development, the Schools Technology Curriculum and the Mathematics decadal plan

Coordination with other sector groups:
- Work with the ACS on curriculum and accreditation
- Work with groups such as ACDICT and government on issues such as CS staff performance metrics and appraisal, and recruitment of students into computing
- A member of CRA (Computing Research Association) and Informatics Europe. These organisations are the North American and European equivalents of CORE.
- A member of Science & Technology Australia, which provides eligibility for Science Meets Parliament and opportunity for input into government policy, and involvement with Science Meets Policymakers

A new Executive Committee from 2013 has been looking at a range of activities that CORE can lead or contribute to, including more developmental activities for CORE members. This has also included a revamp of the mailing lists, creation of discussion forums, identification of key issues for commentary and lobbying, and working with other groups to attract high aptitude students into ICT courses and careers. Again, I welcome your active input into the direction of CORE in order to give our community improved visibility and impact.
CORE’s existence is due to the support of the member departments in Australia and New Zealand, and I thank them for their ongoing contributions, in commitment and in financial support. Finally, I am grateful to all those who gave their time to CORE in 2013, and look forward to the continuing shaping and development of CORE in 2014.

John Grundy
President, CORE
January, 2014
The Australasian Computer Science Week of conferences has been running in some form continuously since 1978. This makes it one of the longest running conferences in computer science. The proceedings of the week have been published as the *Australian Computer Science Communications* since 1979 (with the 1978 proceedings often referred to as Volume 0). Thus the sequence number of the Australasian Computer Science Conference is always one greater than the volume of the Communications. Below is a list of the conferences, their locations and hosts.

2015. Volume 37. Host and Venue - University of Western Sydney, NSW.

2014. **Volume 36. Host and Venue - AUT University, Auckland, New Zealand.**

2013. Volume 35. Host and Venue - University of South Australia, Adelaide, SA.

2012. Volume 34. Host and Venue - RMIT University, Melbourne, VIC.

2011. Volume 33. Host and Venue - Curtin University of Technology, Perth, WA.

2010. Volume 32. Host and Venue - Queensland University of Technology, Brisbane, QLD.


2008. Volume 30. Host and Venue - University of Wollongong, NSW.

2007. Volume 29. Host and Venue - University of Ballarat, VIC. First running of HDKM.

2006. Volume 28. Host and Venue - University of Tasmania, TAS.


1998. Volume 20. Hosts - University of Western Australia, Murdoch University, Edith Cowan University and Curtin University. Venue - Perth, WA.


1995. Volume 17. Hosts - Flinders University, University of Adelaide and University of South Australia. Venue - Glenelg, SA.


1990. Volume 12. Host and Venue - Monash University, Melbourne, VIC. Joined by Database and Information Systems Conference which in 1992 became ADC (which stayed with ACSW) and ACIS (which now operates independently).

1989. Volume 11. Host and Venue - University of Wollongong, NSW.


1987. Volume 9. Host and Venue - Deakin University, VIC.

1986. Volume 8. Host and Venue - Australian National University, Canberra, ACT.


1983. Volume 5. Host and Venue - University of Sydney, NSW.

1982. Volume 4. Host and Venue - University of Western Australia, WA.

1981. Volume 3. Host and Venue - University of Queensland, QLD.

1980. Volume 2. Host and Venue - Australian National University, Canberra, ACT.

1979. Volume 1. Host and Venue - University of Tasmania, TAS.

1978. Volume 0. Host and Venue - University of New South Wales, NSW.
### Conference Acronyms

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Full Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACDC</td>
<td>Australasian Computing Doctoral Consortium</td>
</tr>
<tr>
<td>ACE</td>
<td>Australasian Computing Education Conference</td>
</tr>
<tr>
<td>ACSC</td>
<td>Australasian Computer Science Conference</td>
</tr>
<tr>
<td>ACSW</td>
<td>Australasian Computer Science Week</td>
</tr>
<tr>
<td>ADC</td>
<td>Australasian Database Conference</td>
</tr>
<tr>
<td>AISC</td>
<td>Australasian Information Security Conference</td>
</tr>
<tr>
<td>APCCM</td>
<td>Asia-Pacific Conference on Conceptual Modelling</td>
</tr>
<tr>
<td>AUIC</td>
<td>Australasian User Interface Conference</td>
</tr>
<tr>
<td>AusPDC</td>
<td>Australasian Symposium on Parallel and Distributed Computing (replaces AusGrid)</td>
</tr>
<tr>
<td>AWC</td>
<td>Australasian Web Conference</td>
</tr>
<tr>
<td>CATS</td>
<td>Computing: Australasian Theory Symposium</td>
</tr>
<tr>
<td>HIKM</td>
<td>Australasian Workshop on Health Informatics and Knowledge Management</td>
</tr>
</tbody>
</table>

Note that various name changes have occurred, which have been indicated in the Conference Acronyms sections in respective CRPIT volumes.
ACSW and HIKM 2014 Sponsors

We wish to thank the following sponsors for their contribution towards this conference.

Host Sponsor

Auckland University of Technology,
www.aut.ac.nz

Platinum Sponsor

DATACOM,
www.datacom.com.au

Gold Sponsor

THE UNIVERSITY
OF AUCKLAND
Te Whare Wānanga o Ūtānui Māori
The University of Auckland,
www.auckland.ac.nz

Silver Sponsors

COLAB - AUT Design+Creative Technologies,
colab.aut.ac.nz

Institute of IT Professionals, New Zealand,
www.iitp.org.nz

CORE - Computing Research and Education,
www.core.edu.au

Bronze Sponsor

SERL - AUT Software Engineering Research Laboratory,
www.serl.aut.ac.nz

Australian Computer Society,
www.acs.org.au

Publication Sponsor

University of Western Sydney,
www.uws.edu.au
CONTRIBUTED PAPERS
Crowdsourcing for Clinical Research  An Evaluation of Maturity

Kate E. Birch, Kayla J. Heffernan

No Institute Given
Investigating the Potential of Interactive Media to Encourage Engagement with Type 1 Diabetes Management

Alexander S. Borsboom

No Institute Given
Social Media and Online Survey: Tools for Knowledge Management in Health Research

Mark Merolli, Fernando Jose Martin Sanchez, Kathleen Gray

No Institute Given
BiomRKRS: A Biomarker Retrieval and Knowledge Reasoning System

Bahadorreza Ofoghi, Guillermo Hugo Lopez Campos, Karin Verspoor, Fernando Jose Martin Sanchez

No Institute Given
Automatic Population of Structured Reports from Narrative Pathology Reports

Ying Ou, Jon Patrick

No Institute Given
What Pacific People Think of Online Mental Health Information

Bernadette Peni, Karen Day, Martin Orr

No Institute Given
Health Intelligence: Discovering the Process Model Using Process Mining by Constructing Start-to-End Patient Journeys

Lua Perimal-Lewis, Denise de Vries, Campell H. Thompson

No Institute Given
Developing High Risk Clusters for Chronic Disease Events with Classification Association Rule Mining

Shen Song, Jim Warren, Patricia Riddle

No Institute Given
Building an Empirical Treatment Protocol from High-Resolution Traumatic Brain Injury Data

Anthony Stell, Laura Moss, Ian Piper

No Institute Given
Prescription Patterns for Pacific Patients with Poor Cardiovascular Disease Medication Supply

Jim Warren, Yulong Gu, John Kennelly, Tai Boyce

No Institute Given
Author Index

Birch, Kate E., 3
Borsboom, Alexander S., 13
Boyce, Tai, 89

Day, Karen, 51
de Vries, Denise, 59

Gray, Kathleen, iii, vii, 21
Gu, Yulong, 89

Heffernan, Kayla J., 3
Kennelly, John, 89

Lopez Campos, Guillermo Hugo, 31

Martin Sanchez, Fernando Jose, 21, 31
Merolli, Mark, 21
Moss, Laura, 79

Ofoghi, Bahadorreza, 31
Orr, Martin, 51
Ou, Ying, 41

Patrick, Jon, 41
Peni, Bernadette, 51
Perimal-Lewis, Lua, 59
Piper, Ian, 79

Riddle, Patricia, 69
Song, Shen, 69
Stell, Anthony, 79

Thompson, Campell H., 59
Verspoor, Karin, 31
Warren, Jim, iii, vii, 69, 89
Listed below are some of the latest volumes published in the ACS Series Conferences in Research and Practice in Information Technology. The full text of most papers (in either PDF or Postscript format) is available at the series website http://crpit.com.

**Volume 124 - Database Technologies 2012**
- Contains the proceedings of the Twenty-Third Australasian Database Conference (ADC 2012), Melbourne, Australia, 30 January – 2 February 2012.

**Volume 125 - Information Security 2012**
- Contains the proceedings of the Tenth Australian Information Security Conference (AISC 2012), Melbourne, Australia, 30 January – 3 February 2012.

**Volume 126 - User Interfaces 2012**
- Edited by Baifeng Chen, Flinders University, Australia and Rose T. Smith, University of South Australia, Australia. January 2012. 978-1-921770-07-1.
- Contains the proceedings of the Thirteenth Australian User Interface Conference (AUIC2012), Melbourne, Australia, 30 January – 3 February 2012.

**Volume 127 - Parallel and Distributed Computing 2012**
- Contains the proceedings of the Tenth Australasian Symposium on Parallel and Distributed Computing (AusPDC 2012), Melbourne, Australia, 30 January – 3 February 2012.

**Volume 128 - Theory of Computing 2012**
- Edited by Julián Mestre, University of Sydney, Australia. January 2012. 978-1-921770-09-5.
- Contains the proceedings of the Eighteenth Computing: The Australasian Theory Symposium (CATE 2012), Melbourne, Australia, 30 January – 3 February 2012.

**Volume 129 - Health Informatics and Knowledge Management 2012**
- Contains the proceedings of the Eighth Asia-Pacific Conference on Conceptual Modelling (APCCM 2012), Melbourne, Australia, 31 January – 3 February 2012.

**Volume 130 - Conceptual Modelling 2012**

**Volume 133 - Australian System Safety Conference 2011**

**Volume 134 - Data Mining and Analytics 2012**
- Edited by Xiong Zeng, Department of Immigration and Citizenship, Australia, Junying Li, University of South Australia, Australia, Paul J. Kennedy, University of Technology, Sydney, Australia and Peter Christen, Australian National University, Australia. December 2012. 978-1-921770-14-2.
- Contains the proceedings of the Tenth Australasian Data Mining Conference (AusDM’12), Sydney, Australia, 5-7 December 2012.

**Volume 135 - Computer Science 2013**
- Edited by Bruce Thomae, University of South Australia, Australia. January 2013. 978-1-921770-20-3.
- Contains the proceedings of the Thirty-Sixth Australasian Computer Science Conference (ACSC 2013), Adelaide, Australia, 29 January – 1 February 2013.

**Volume 136 - Computing Education 2013**
- Edited by Angela Carbone, Monash University, Australia and Jacqueline Whalley, AUT University, New Zealand. January 2013. 978-1-921770-21-0.

**Volume 137 - Database Technologies 2013**
- Contains the proceedings of the Twenty-Fourth Australasian Database Conference (ADC 2013), Adelaide, Australia, 29 January – 1 February 2013.

**Volume 138 - Information Security 2013**

**Volume 139 - User Interfaces 2013**
- Edited by Rose T. Smith, University of South Australia, Australia and Burkhard C. Wünsche, University of Auckland, New Zealand. January 2013. 978-1-921770-26-5.
- Contains the proceedings of the Fourteenth Australasian User Interface Conference (AUIC 2013), Adelaide, Australia, 29 January – 1 February 2013.

**Volume 140 - Parallel and Distributed Computing 2013**
- Contains the proceedings of the Eleventh Australasian Symposium on Parallel and Distributed Computing (AusPDC 2013), Adelaide, Australia, 29 January – 1 February 2013.

**Volume 141 - Theory of Computing 2013**

**Volume 142 - Health Informatics and Knowledge Management 2013**
- Contains the proceedings of the Sixth Australasian Workshop on Health Informatics and Knowledge Management (HIKM 2013), Adelaide, Australia, 29 January – 1 February 2013.

**Volume 143 - Conceptual Modelling 2013**
- Contains the proceedings of the Ninth Asia-Pacific Conference on Conceptual Modelling (APCCM 2013), Adelaide, Australia, 29 January – 1 February 2013.

**Volume 144 - The Web 2013**
- Contains the proceedings of the First Australian Web Conference (AWC 2013), Adelaide, Australia, 29 January – 1 February 2013.

**Volume 145 - Australian System Safety Conference 2012**
- Contains the proceedings of the Australian System Safety Conference (ASSC 2012), Brisbane, Australia, 23rd – 25th May 2012.
Contributed Papers
Crowdsourcing for Clinical Research – An Evaluation of Maturity

KATE E. BIRCH\textsuperscript{ab} and KAYLA J. HEFFERNAN\textsuperscript{b}

\textsuperscript{a} Melbourne Brain Centre @ Royal Melbourne Hospital, Department of Medicine, University of Melbourne  
\textsuperscript{b} Computing and Information Systems, Melbourne School of Engineering, University of Melbourne

kate.birch@unimelb.edu.au kaylah@student.unimelb.edu.au

Abstract

With the growth of the Internet and individuals using the Internet for person health research, crowdsourcing clinical research has the potential to become a powerful tool in surveilling and monitoring health outcomes. This paper evaluates the maturity of the emerging tool of crowdsourcing clinical research using two carefully selected and adapted evaluation models: Project Management Maturity Model (ProMMM) and National Infrastructure Maturity Model (NIMM). Two models were used in conjunction for the evaluation as ProMMM focuses on a professional’s ability to utilise crowdsourcing for clinical research, while NIMM focuses on the maturity of crowdsourcing clinical research itself. To evaluate maturity, the authors reviewed available literature and conducted primary research in the form of interviews at the Melbourne Brain Centre at Royal Melbourne Hospital with Associate Professor Helmut Butzkueven, MS Neurologist and Researcher, and Dr Athina (Tina) Soulis, General Manager of Neuroscience Trials Australia. The tool of crowdsourcing for clinical research and the users and prospective users of the tool were found to be in immaturity. Despite immaturity, the future holds exciting applications for crowdsourcing clinical research with the potential to save costs, time, and recruit wider cohorts into clinical research.

Keywords: crowdsourcing; clinical research; maturity; evaluation.

1. Introduction

Crowdsourcing is a popular method of obtaining services, ideas, designs and even funds by putting out an “open call” for contributions. Swan (2012) defines crowdsourcing as: “the practice of obtaining participants, services, ideas, or content by soliciting contributions from a large group of people, especially via the Internet”. This concept is not new; an early example of crowdsourcing is the Longitude Prize, an award offered by the British government in 1714 for the invention of a simple way of determining a ship’s longitude (Sobel, 2004). The emergence of web 2.0 has greatly increased the capacity and popularity of crowdsourcing (Swan, 2012). A PEW Research Center study published in 2011 (Fox) found that of the adults surveyed:

- 74% use the Internet;
- 80% of Internet users have looked online for health information;
- 34% of Internet users have read someone else’s commentary or experience about health or medical issues; and
- 18% of Internet users have gone online to find others who have health concerns similar to theirs.

This high usage of the Internet for personal health research supports the notion that crowdsourcing for clinical research could be a powerful tool. With the addition of citizen science, there has been a recent rise in coordinated self-experimentation, leading to participant-organised (Roberts, 2011; Roberts, 2010; Martin, Burns, and Doiron, 2011) and researcher initiated (Turner et al., 2011; Wicks et al., 2011a; Wicks et al., 2009; Frost et al., 2011; Wicks et al., 2011b) studies. Researchers are now sourcing entire cohorts and collecting data directly from social media sources.

This evaluation seeks to determine the maturity of crowdsourcing for clinical research by examining published work utilising crowdsourcing for cohort definition or data collection. AB et al. (2013) identifies four main types of crowdsourced research; problem solving; data processing; surveillance/monitoring; and surveying. This paper will predominantly focus on crowdsourced surveillance/monitoring.

The aim of this evaluation is to determine the maturity of the method of crowdsourced clinical research, with focus on evidence for the tool over level of adoption as a primary indicator, with an outcome focus instead of a return on investment model.

2 Background

Crowdsourcing clinical research enables researchers to recruit participants and obtain data from large numbers of patients via the Internet. Crowdsourcing affords researchers a new way to survey diseases “especially given the recent ubiquity of information technology tools that can automate and accelerate the data collection process” over communities in which patients are already volunteering this information (Chunara, 2012), offering a continual pool of participants (Turner, Kirchhoff and Capurro, 2012).

The intended benefits of utilising crowdsourcing for clinical research include cost savings, reduction of geographical limitations and shorter recruitment periods than traditional clinical research. Additional advantages outlined by Freifeld et al. (2010) include scalability, coverage, timeliness, and transparency. Use of crowdsourcing for recruitment substantially reduces the
costs. A recent study found the cost of such recruitment was $4.82US per patient, compared with $86.28US for direct mail and $195.65US for email enlisted participants (Cascade et al., 2012). Overcoming such costs may enable researchers to increase trial sizes and afford them the ability to enlist participants “in more diverse geographic regions, without incurring the costs of dedicated oversight teams in multiple locations” (Cook, 2011). With many thousands of patients already sharing their health data online, researchers can readily identify suitable candidates for recruitment to specific studies, expediting the entire research process, as recruitment accounts for 45% of study delays (Cascade et al., 2012).

Tepper (2013) quotes Dr Chad Cook, a researcher who has used this technique (his work was reviewed as part of the evaluation), as saying crowdsourcing in this context is “examining the mundane to see if it would affect clinical practice. We can investigate questions with a larger pool of patients and clinicians, and to answer questions we normally couldn’t answer”.

One of the most prominent studies to date utilizing crowdsourcing in an interventional setting is the investigation of the experimental use of lithium carbonate treatment to treat amyotrophic lateral sclerosis (ALS). ALS is a rapidly neurodegenerative disease that causes weakness and atrophy, with a median survival from symptom onset of two (2) to five (5) years (Wijesekera and Leigh, 2009). A small, single blinded study showed that lithium carbonate treatment was potentially efficacious in the treatment of ALS (Fornai et al., 2008). Despite clinician scepticism, patients were enthusiastic to try this treatment, and began taking the medication off-label. Patients used an online spread-sheet to gather data, and PatientsLikeMe®, an online network that enables people with similar conditions to connect, (PatientsLikeMe®, 2013) built a lithium-specific data collection tool for the use of their 348 lithium-taking ALS patients. Wicks et al. (2011b) analysed the data collected, and attempted to overcome potential biases created in the absence of randomisation, blinding and placebo control. The researchers used a matching algorithm to match 148 eligible-for-analysis participants to 447 controls using historical data collected on PatientsLikeMe®. Lithium carbonate treatment was found to have no effect of ALS disease progression. The researchers acknowledged this study method is not a substitute for double blind randomised control trials (RCT), but the outcome of this study was later replicated in RCTs (Gamez, Salvado and Badia, 2013).

### 3 Methodology

According to van de Wetering and Batenburg (2009) theories of information systems and IT maturity and adoption are well established, dating back to the early 1970’s. The first known model is the Nolan Model, which was the basis for the evaluation model search. An initial literature search was conducted using Google Scholar and University of Melbourne Discovery. Several articles discussing evaluation frameworks were identified; van de Wetering and Batenburg (2009); Persson and Goldkuhl (2005); O’Neil (2011); and Galliers and Sutherland (1991). From these articles, the frameworks presented in Table 1 were identified as being potentially appropriate to evaluate crowdsourcing clinical research.

<table>
<thead>
<tr>
<th>The Nolan Model</th>
<th>Project Management Maturity Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Initiation</td>
<td>• Naïve</td>
</tr>
<tr>
<td>• Contagion</td>
<td>• Novice</td>
</tr>
<tr>
<td>• Control</td>
<td>• Normalised</td>
</tr>
<tr>
<td>• Integration</td>
<td>• Natural</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Stages of Project Management</th>
<th>Lee and Layne Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Paulk et al., 1993)</td>
<td>(Layne and Lee, 2001)</td>
</tr>
<tr>
<td>• Performed</td>
<td>• Catalogue</td>
</tr>
<tr>
<td>• Managed</td>
<td>• Transaction</td>
</tr>
<tr>
<td>• Defined</td>
<td>• Vertical Integration</td>
</tr>
<tr>
<td>• Quantitatively Managed</td>
<td>• Horizontal Integration</td>
</tr>
<tr>
<td>• Optimising</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Constitutes and the e-diamond Model</th>
<th>e-readiness Categories</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Albinsson et al., 2006)</td>
<td>(van Dyk, Schutte and Fortuin, 2012)</td>
</tr>
<tr>
<td>• Separated</td>
<td>• Technology / Maintenance</td>
</tr>
<tr>
<td>• Coordinated</td>
<td>• Policy and Legal</td>
</tr>
<tr>
<td>• General</td>
<td>• Individual Users</td>
</tr>
<tr>
<td>• Individual</td>
<td>• Organizational Processes</td>
</tr>
<tr>
<td>• Information</td>
<td>• Planning</td>
</tr>
<tr>
<td>• Quantitatively Managed</td>
<td>• Financial Sustainability</td>
</tr>
<tr>
<td>• Optimising</td>
<td>• Interaction / Community Involvement</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Layered Telemedicine Implementation</th>
<th>The Seven S’s (Pascale and Athos, 1981)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Broens et al., 2007)</td>
<td></td>
</tr>
<tr>
<td>• Prototype</td>
<td>• Strategy</td>
</tr>
<tr>
<td>• Small-scale Pilots (Acceptance)</td>
<td>• Structure</td>
</tr>
<tr>
<td>• Large Scale Pilots (Financing, Org)</td>
<td>• Systems</td>
</tr>
<tr>
<td>• Operational (Policy and Legislation)</td>
<td>• Staff</td>
</tr>
<tr>
<td>• Performance (Financial, Organizational)</td>
<td>• Style</td>
</tr>
<tr>
<td></td>
<td>• Skills</td>
</tr>
<tr>
<td></td>
<td>• Superordinate goals</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>The Hirschheim et al. Model</th>
<th>National Infrastructure Maturity Model (NHS, 2011)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Hirschheim et al., 1988)</td>
<td></td>
</tr>
<tr>
<td>• Delivery</td>
<td>• Level 1: Initial, ad hoc, process (basic)</td>
</tr>
<tr>
<td>• Reorientation</td>
<td>• Level 2: Managed, stable process (controlled)</td>
</tr>
<tr>
<td>• Reorganisation</td>
<td>• Level 3: Defined, standard (standardized)</td>
</tr>
</tbody>
</table>
Table 1: Potential Frameworks to Evaluate Crowdsourcing Clinical Research

The authors then reviewed this table and eliminated models that were not relevant in the clinical research methodological space, leaving five (5). These were further examined against the following criteria:

1. Framework evaluates maturity;
2. Framework focuses on evidence for the tool over level of adoption as a primary indicator of maturity; and
3. Framework is outcome focused, not Return on Investment focused.

The results of this analysis are presented in Table 2.

<table>
<thead>
<tr>
<th>Criteria #</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Nolan Model</td>
<td>✓</td>
<td>✗</td>
<td>✓</td>
</tr>
<tr>
<td>Stages of Project Management Maturity</td>
<td>✓</td>
<td>✗</td>
<td>✓</td>
</tr>
<tr>
<td>PROMMM</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>NIMM</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>E-readiness categories</td>
<td>✓</td>
<td>✗</td>
<td>✓</td>
</tr>
</tbody>
</table>

Table 2: Analysis of Evaluation Framework Against Criteria

Two evaluation models met all three criteria: Project Management Maturity Model (ProMMM) (Hillson, 2003) and the National Infrastructure Maturity Model (NIMM) (van Dyk, Schutte and Fortuin, 2012). The authors determined the use of both evaluation measures to be appropriate as ProMMM focuses on the ability of professionals to utilise the tool while NIMM focuses on the maturity of the tool itself.

ProMMM (Hillson, 2003) is designed to assess the level of capability of project managers and is often applied to organisations to determine how mature they are. In the context of this paper, stages relate to healthcare professionals at an individual level. ProMMM has the following levels of maturity:

- Naïve;
- Novice;
- Normalised; and
- Natural.

To assess the maturity of crowdsourcing clinical research using ProMMM, the stage-characteristic model in Table 3 was used, adapted from Hillson (2003).

<table>
<thead>
<tr>
<th>Level</th>
<th>Characteristics of Stage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Naïve</td>
<td>• Potential users unaware of the value of the tool&lt;br&gt; • No structured approach to use&lt;br&gt; • Culture is resistant to change&lt;br&gt; • Need for tool recognised&lt;br&gt; • No experience of use&lt;br&gt; • No application for the tool</td>
</tr>
<tr>
<td>Novice</td>
<td>• Small number of users have begun to experiment with tool&lt;br&gt; • No formal or structure of generic processes</td>
</tr>
<tr>
<td>Normalised</td>
<td>• Implemented across organisation / industry&lt;br&gt; • Formalised&lt;br&gt; • Recognised value of tool&lt;br&gt; • Understand expected benefits of tool&lt;br&gt; • Users have experience and expertise&lt;br&gt; • Application of tool is routine and consistent</td>
</tr>
<tr>
<td>Natural</td>
<td>• Accepted culture across organisation / industry&lt;br&gt; • Best-practice usage&lt;br&gt; • All potential users have a degree of experience&lt;br&gt; • Application is widespread and second-nature</td>
</tr>
</tbody>
</table>

Table 3: PRoMMM Maturity Level Characteristics (Hillson, 2003)

The NHS developed a National Infrastructure Maturity Model (NIMM) to assess the IT Infrastructure of the UK National Health Service (van Dyk, Schutte and Fortuin, 2012). In the context of this paper, crowdsourcing clinical research will be assessed against the levels to determine the maturity of the tool. These levels are:

- Level 1: Initial, ad hoc process (Basic);
- Level 2: Managed, stable process (Controlled);
- Level 3: Defined, standard process (Standardised);
- Level 4: Measured process (Optimised); and
- Level 5: Optimizing (Innovative).

To assess the maturity of Crowdsourcing clinical research using NIMM, the stage-characteristic model in Table 4 was used, adapted from Essmann (2009).
Table 4: NIMM Maturity Level Characteristics (NHS, 2011)

<table>
<thead>
<tr>
<th>Level</th>
<th>Characteristics of Stage</th>
</tr>
</thead>
</table>
| 1: Initial, ad hoc process (Basic) | • Ad hoc and chaotic usage  
| | • Used by individuals only |
| 2: Managed, stable process (Controlled) | • Use of tool planned, performed, measured and controlled  
| | • Documented use  
| | • Requirements, processes of tool are managed  
| | • Commitments are established |
| 3: Defined, standard process (Standardised) | • The tool is well characterized and understood  
| | • Standards, procedures and methods for tool use  
| | • Consistent usage  
| | • More rigour in use |
| 4: Measured process (Optimised) | • Quality and process performance of tool use is understood in statistical terms  
| | • Detailed measures of tool performance |
| 5: Optimizing process (Innovative) | • Usage continually improved based on a quantitative understanding  
| | • Focus is on continually improving tool performance  
| | • Shared learning |

4.1 Evidence for Naïve Level

4.1.1 Potential Users Unaware of the Value of the Tool – Substantiated, but Reducing Unawareness

Several review and opinion articles have appeared in peer-reviewed journals (AB et al., 2013; Behrend et al., 2011; Cascade et al., 2012; Cheeney, Hariskamp and Schupp, 2012; Cook, 2011; Ekins and Williams, 2010; Norman et al., 2011; Swan, 2012; Tepper, 2013) expounding the benefits, limitations and current landscape of crowdsourcing clinical research. The literature predominantly discusses crowdsourcing as a new area, and thus while not all potential users are aware of benefits, there is knowledge transfer growth in this area.

4.1.2 There is no Structured Approach to Use - Substantiated

There are no formal standards for use and application is neither routine nor consistent. AB et al. (2010) found “considerable variability in how the methods of crowdsourcing were reported”. Crowdsourced clinical research is not accepted industry-wide and “do not always conform to generally accepted industry practices of research conduct” (Swan, 2012).

4.1.3 Culture is Resistant to Change – Substantiated

Swan (2012) addresses the belief that some in the medical research field are reported to have, that “citizen science in not really science”. The limiting belief that science is an esoteric exercise to be practiced by a few gifted minds may be overcome through the early successes of crowdsourcing. Results from crowdsourced studies have been published in peer-reviewed journals, but more often in grey literature (Swan, 2012). This may reflect the journal editors’ belief that there is lack of acceptance of the tool in the broader scientific community.

Tepper (2013) states one of the major sources of reluctance in the uptake of crowdsourcing in clinical research is the low quality of data generated and the need to filter information adding to time and cost requirements. Swerlick, James and Minnillo (2011) argue this limitation is actually an advantage that will attract researchers. They describe the physician as a “filter” in the data collection process, which results in missing drug induced adverse events.

4.1.4 Need for Tool is Not Recognised – Unsubstantiated

The need to generate cost savings, reduce geographical limitations of participation and decrease recruitment periods is well established. Crowdsourcing has been found to reduce recruitment costs (Cascade et al., 2012), is immune to geographical restrictions, and has a large pool of potential participants at the ready. Crowdsourcing also offers the advantages of scalability, coverage, timeliness, and transparency (Freifeld et al., 2010).

The authors determined the maturity level of crowdsourcing for clinical trials using both models, based on evidence for each stage in available literature and primary interviews conducted at the Melbourne Brain Centre at Royal Melbourne Hospital with Associate Professor Helmut Butzkueven, MS Neurologist and Researcher, and Dr Athina (Tina) Soulis, General Manager of Neuroscience Trials Australia.

4 Results – ProMMM Maturity Model

Initial readings of current crowdsourcing literature indicate it is new to clinical research, and so it is expected that users and prospective users will fall into the Naive or Novice categories.
4.1.5 No Experience of Use - Unsubstantiated

This criterion is not met as there is published experience of use; for example in Wicks et al. (2011b); Turner et al. (2011); Wicks and MacPhee (2009); Frost et al (2011); Cheeney, Harskamp and Schupp (2012); and AB et al. (2013).

4.1.6 No Application for the Tool – Unsubstantiated

The for-profit sector is enthusiastic about possibilities offered by crowdsourcing. Norman et al. (2011) quotes Dr. Paul Chapman, Head of Takeda’s Pharmaceutical Research Division as saying: “In particular I was very excited about the opportunities generated by crowdsourcing clinical research. In neuroscience, for example, we are often faced with very difficult decisions about which of several unmet medical needs to be addressed first, with a good compound for a novel target. Precompetitive crowdsourcing would mean that we could get multiple shots on a competitive compound, we can understand how to align our compound with the most promising unmet medical need”.

A/Prof Butzkueven (2013) raises another application of the tool: “we have trouble in terms of conducting clinical trials in Australia, recruitment is going down and these sorts of communication strategies enable us to reach lots more people at once”. Dr Soulis (2013) adds, “in the 20 years I’ve been involved in clinical trials I think I can count on one hand the number of trials that recruited on time or faster than expected. Recruitment is always an issue. This is a really good tool or vehicle to address this area of recruitment; that would be its biggest advantage”.

4.2 Evidence for Novice Level

4.2.1 Small Number of Users Have Begun to Experiment with Tool – Substantiated

Frost et al. (2011) acknowledge use of “patient-reported outcomes entered via an online community” is a new source of evidence to evaluate off-label medication use. They expound the benefits of crowdsourcing data in their discussion, which is a practice unique to new and experimental research methods. Frost et al. (2011) explain their crowdsourcing method was able to collect previously unrecorded data types from large and increasingly diverse populations. Similarly, Wicks et al. (2011a) describe online communities as “an opportunity” that has not yet been fully explored.

AB et al. (2013) states “crowdsourcing clearly is not used pervasively in health research”. Evidence for more experimentation with crowdsourcing clinical research is accepted, the ethical principles underpinning the tool are well established. Drug discovery and safety are well-established areas of research requirement. Wicks et al. (2011b) discuss obligations of researchers to collect data regarding the safety of self-experimentation, and of potentially efficacious drugs discovered in crowdsourced data.

Cook (2011) believe’s that while widespread adoption has not yet occurred, “this is a research model of the future. And as far as having clinicians drive the success of dedicated research projects, I cannot think of a better group to do it”. Tippler (2013) states the implementation of this tool will create economic disruptions, as funding models are not currently adaptive to this tool.

4.3 ProM MMM Results Summary

<table>
<thead>
<tr>
<th>Level</th>
<th>Characteristics of Stage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Naïve</td>
<td>✓ Potential users unaware of the value of the tool</td>
</tr>
<tr>
<td></td>
<td>✓ No structured approach to use</td>
</tr>
<tr>
<td></td>
<td>✓ Culture is resistant to change</td>
</tr>
<tr>
<td></td>
<td>✗ Need for tool recognised</td>
</tr>
<tr>
<td></td>
<td>✗ No experience of use</td>
</tr>
<tr>
<td></td>
<td>✗ No application for the tool</td>
</tr>
<tr>
<td>Novice</td>
<td>✓ Small number of users have begun to experiment with tool</td>
</tr>
<tr>
<td></td>
<td>✓ No formal or structure generic processes</td>
</tr>
<tr>
<td>Normalised</td>
<td>✗ Implemented across organisation / industry</td>
</tr>
<tr>
<td></td>
<td>✗ Formalised</td>
</tr>
<tr>
<td></td>
<td>✗ Recognised value of tool</td>
</tr>
<tr>
<td></td>
<td>✗ Understanding expected benefits of tool</td>
</tr>
<tr>
<td></td>
<td>✗ Users have experience and expertise</td>
</tr>
<tr>
<td></td>
<td>✗ Application of tool is routine and consistent</td>
</tr>
<tr>
<td>Natural</td>
<td>✗ Accepted culture across organisation / industry</td>
</tr>
</tbody>
</table>
Practitioners using crowdsourcing for clinical research fall in the Novice level of maturity according to the ProMmM maturity model. A/Prof Butzkueven (2013) confirms this outcome: “we are just at the starting point”.

5 Results NIMM Maturity Model

As crowdsourcing for clinical research is a new tool it is expected to fall into the Level 1, Basic, or Level 2, Controlled, categories of NIMM.

5.1 Evidence for Level 1 - Basic

5.1.1 Ad Hoc and Chaotic Usage - Substantiated

Crowdsourcing for clinical research is well published, including but not limited to papers by Wicks et al. (2011b); Turner et al. (2011); Wicks and MacPhee (2009); Frost et al. (2011a); Cheeney, Harskamp and Schupp, (2012); and AB et al. (2013). However, usage is currently limited to a few pockets of research interests.

Swerlick, James and Minnillo (2011) describe the crowdsourcing tool as “a green field where the options for what and where to build are wide open” and that this tool could be used in “an almost unlimited number of clinically relevant studies” (Swerlick, James and Minnillo, 2011).

5.1.2 Used by Individuals Only - Substantiated

Swan (2012) states: “thus far the only form of study conducted by professional researchers in crowdsourced cohorts has been retrospective, non-intervention user questionnaires”. The tool needs further research and refinement to be applicable to a wider range of research questions and protocols (Swan, 2012). Once this development has occurred, institutions and companies may adopt the technique more widely, but no evidence was found that the tool is being used across departments, institutions or companies.

5.2 Evidence for Level 2 - Controlled

5.2.1 Use of Tool Planned, Performed, Measured and Controlled - Unsubstantiated

Cheeney, Harskamp and Schupp (2012) note the crowdsourcing technique is currently limited by lack of standardised data collection and validated instruments. Armstrong et al. (2012) echoes the call for validation of measures used in crowdsourced data collection. As does Swan (2012): “crowdsourced health research studies do not always follow the rigorous protocols”. Cheeney, Harskamp and Schupp (2012) also discuss the lack of convergence in evidence surrounding this tool. For example, some studies have shown that crowdsourcing reaches an older, more ethnically diverse population (Behrend, 2011), while other have shown crowdsourcing populations can be more defined and less representative in demographic makeup.

Armstrong et al. (2012) reports the quality of data obtained through crowdsourcing is unknown, but Behrend (2011) showed data they accumulated through a crowdsourcing effort was as good or better than data collected in a corresponding university sample. Cheeney, Harskamp and Schupp (2012) found dramatic differences in the rates of efficacy of acne treatments. What is unknown is if this is the result of low quality data being collected using crowdsourcing, or a true result that would not otherwise be found due to the difference in cohorts, clinician filtering and research design. A/Prof Butzkueven (2013) provides an example of the insufficient measurement and control of the tool; “for example, you have a disease specific website where people can register and then record their medications, and can report patient reported outcome...you could theoretically use that data for an efficacy study for different medications but you end up with lots of selection biases. People lose interest or become too unwell will stop posting or engaging. They are really just ballooned selection biases that we get in registered studies anyway. So like any other scientific endeavour, the methodologies have to be examined and it has to be peer reviewed”.

Frost et al. (2011), Cheeney, Harskamp and Schupp (2012), Swan (2012), Wicks et al. (2011b) and AB et al. (2013) all address the limitations of crowdsourcing, resultant from a lack of research in to, and refinement, of the tool. The effect of patient motivations in: reporting self-experimentations; the lack of demographic data collection; the temporal habits of online self-report; the lack of diagnosis confirmation; the possible over-reporting of positive outcomes; and, the possibility that people are deliberately misrepresenting themselves have not yet been adequately examined. Dr Soulis (2013) adds to this, saying: “it’s ploughing through the information that would be available... how do you substantiate the validity of the information... how do you work that out? That is a major limitation”.

5.2.2 Documented Use – Substantiated

As previously mentioned, there is documented use of the tool, including but not limited to: Wicks et al. (2011a); Turner et al., (2011); Wicks and MacPhee (2009); Frost, et al. (2011); Cheeney, Harskamp and Schupp (2012); and AB et al. (2013).

5.2.3 Requirements, Processes of Tool Use are Managed – Unsubstantiated

Tepper (2013) discusses that crowdsourcing techniques are “outrunning legal and regulatory protections... we are in the infancy of case law related to the privacy of social network health information”. Swan (2012) elaborates by stating that crowdsourced clinical research do not
conform to standard and accepted industry practices. In fact, studies specifically point out traditional compliance methods are not practicable, and institutional review boards may not approve crowdsourced research for this reason. Dr Soulis (2013) states that “the regularly environment is not ready for it. In the global sense there would need to be guidance from the big regulators like the FDA, EMA in Europe, like our TGA... Regulatory agencies are really keen on proving that you have done your trial in the best faith... making sure they (participants) fully understand what they are signing up for”.

A/Prof Butzkueven (2013) reports that there is a move toward acceptance of these methods: “the ethics committees that regulate research seem to be on board with this in general, but of course there are some ethical limitations”.

5.2.4 Commitments are Established - Substantiated
A commitment to use has been made by the pharmaceutical industry as they recognise the value of crowdsourcing clinical research “to reduce research and development costs” (Ekins and Williams, 2010).

The ethical principles underpinning the tool are well established. Drug discovery and safety are two well-established areas of research need.

5.2.5 NIMM Results Summary

<table>
<thead>
<tr>
<th>Level</th>
<th>Characteristics of Stage</th>
</tr>
</thead>
</table>
| 1:Initial, ad hoc process (Basic) | ✓ Ad hoc, chaotic usage  
✓ Used by individuals |
| 2: Managed, stable process (Controlled) | ✓ Use of tool planned, performed, measured and controlled  
✓ Documented use  
✗ Requirements and processes of tool are managed  
✓ Commitments are established |
| 3:Defined, standard process (Standardised) | ✓ Well characterized and understood  
✗ Stands, procedures and methods for tool use  
✗ Consistent usage  
✗ More rigour in use |
| 4:Measured process (Optimised) | ✓ Quality and process performance of tool use is understood in statistical terms  
✗ Detailed measures of tool performance  
✗ Usage continually improved based on a quantitative understanding  
✗ Focus is on continually improving tool performance  
✗ Shared learning |

Table 6: Summary of Results – Evaluation using NIMM

Using the NIMM Maturity Model to look at the maturity of crowdsourcing clinical research as a tool places it at Level 1: Basic, initial, ad hoc process.

6 Conclusions and Recommendations
This evaluation demonstrates crowdsourcing clinical research is a relatively immature tool and is sparsely used by both individuals (ProMMM evaluation), and in the field of health research on a whole (NIMM evaluation).

Low maturity tools are often riskier and harder to operate, as users are unfamiliar with the application. In a highly regulated field, such as health, use of immature tools is especially difficult to justify as focus is on evidence-based methods of research for informing health actions.

With the current limitations of crowdsourcing clinical research: self-selection bias; funding limitations; and, shortcomings in study design (Swan, 2012), crowdsourcing needs further research and refinement for it to be applicable to a wider range of research questions and protocols, and to be competitive with the ‘gold standard’ of RCT’s.

The authors acknowledge further limitations including demographic disparity between groups that do and do not use the Internet or participate in online health communities. For example, “women, non-Hispanic whites, younger adults, and those with higher levels of education and income” are more likely to use the Internet to gather health information than other demographic groups (Fox, 2011). This paper determined the maturity level of crowdsourcing clinical research at the individual professional and industry levels, however the authors identify areas for future research exist to evaluate benefits and limitations of crowdsourcing clinical research, including cost-benefit analysis and validation studies.

Despite such limitations the future holds exciting applications for crowdsourcing clinical research, including: demographic collection (Cheeney, Harskamp and Schupp, 2012); (Armstrong et al., 2012); images for patients to report against (Cheeney, Harskamp and Schupp, 2012)(Armstrong et al., 2012); treatment periods and medication adherence (Cheeney, Harskamp and Schupp, 2012)(Armstrong et al., 2012); and, controls of reporting of new signs and symptoms with clinical review.
to prevent duplication (Cheeney, Harskamp and Schupp, 2012)(Armstrong et al., 2012). With more people using the Internet, crowdsourcing clinical research is an exciting tool that can allow clinical researchers the ability to harness currently untapped data, perhaps in conjunction with traditional methods. As Swerlick, James and Minnillo (2011) summarises: “the possibilities are endless”, but at this stage, crowdsourcing is just that, a possibility, and not a proven mechanism.

7 References


Butzkueven, H. (2013): Interview by authors at Melbourne Brain Centre Royal Melbourne Hospital, Wednesday 24th July 2013.


Norman, T.C., Bountra, C., Edwards, A.M., Yamamoto,


Soulis, A. (2013): Interview by authors at Melbourne Brain Centre Royal Melbourne Hospital, Thursday 25th July 2013.


Investigating the potential of Interactive Media to Encourage Engagement with Type 1 Diabetes Management

ALEXANDER S. BORSBOOM1

1 Department of Software Engineering
The University of Auckland,
Private Bag 92019, Auckland, New Zealand,
Email: alexander.borsboom@gmail.com

Abstract

The effective management of type 1 diabetes is very challenging for a number of people. Improving the management skills of those suffering with the disease is even more so due to the sheer number of Type 1 diabetes sufferers. We believe that a serious video game which targets diabetes education has the potential to significantly improve the management of Type 1 Diabetes. In light of this we have developed a series of game mechanics which are able to represent Type 1 Diabetes management in a video game. We have developed a prototype video game to demonstrate the effectiveness of these game mechanics, and we have identified potential areas of future development in this field.

Keywords: Video Game; Serious Video Game; Diabetes; Medical Video Games

1 Introduction

Diabetes is a serious medical condition with far reaching consequences on the lives of those living with the disease and their friends and family. Effective management of diabetes requires the management of many interrelating systems such as diet management, exercise management and medication management (Hanas 2007). Failure to effectively manage the symptoms of diabetes can lead to serious health consequences: Hyperglycemic episodes can lead to blindness and organ failure, while hypoglycemic episodes can lead to fainting, mood swings, anxiety and dizziness. Effective management of diabetes symptoms is challenging and an area of interest to many. Better symptom management has been shown to reduce the long term health impact of diabetes on patients lives (Norris, Lau, Smith, Schmid & Engelgau 2002)(Norris, Engelgau & Narayan 2001) (Deakin, McShane, Cade & Williams 2005), as well as reducing the financial burden the disease places upon New Zealand as a whole (Wagner, Sandhu, Newton, McCulloch, Ramsey & Grothaus 2001). Additionally, it is known that tailored healthcare which accurately represents and deals with the issues facing individuals and their circumstances is significantly more effective at improving healthcare outcomes when compared to non-tailored solutions (Mensing & Norris 2003) (Rickheim, Weaver, Flader & Kendall 2002)

(Hiss 1996). Unfortunately, Diabetes is very common and thus the large number of sufferers to whom assistance must be provided directly impacts the viability of many proposed diabetes treatment plans. One solution which remains unhindered by the volume of diabetes patients is video games. A video game can be replicated many times and thus is suitable for distribution amongst large numbers of people. Video games can also be designed in such a way that they are able to dynamically reconstruct content in response to a number of user specified properties, thus allowing the gameplay experience to change to represent the circumstances and specific issues facing a player (Booth 2009). This report details the construction of a prototype video game “proof of concept” which demonstrates the feasibility of creating a serious video game to teach people about effective diabetes management. The report also outlines a set of game mechanics which can be used as an effective representation of Type 1 Diabetes in a video game. The video game proof of concept was produced using the Unity 3D engine.

2 Serious Video Games, Educational Games and Gamification

Traditionally, a video game consists of a rule based system with which the user interacts for their own enjoyment. These rules which make up the core video game interactions are referred to as game mechanics. The focus of the video game is entirely placed upon the user’s enjoyment, which usually is achieved by overcoming challenges, although more recently some video games have achieved this by allowing the player some form of self expression, such as the extremely popular video game Minecraft, in which players build their own worlds, buildings, mining systems and industries.

Educational video games have existed since the advent of video games themselves. We use the term “Educational Game” here to refer to a specific style of video game common to early educational video games and children’s learning games. We define the term as a game where the rule based system which forms the core of the video game relates to knowledge. For example, there is geographic knowledge in the core of the popular Where in the world is Carmen San Diego? series, or there is skill based knowledge, such as algebra or mathematical skills, in other children’s games. Users must complete challenges using these skills, thus the game reinforces gained knowledge and encourages education. This form of video game is very common in the educational space ( sometimes referred to as an edutainment game ) and is distinct from other genres of video game. When dealing with diabetes, it is very challenging to accurately express the informa-
tion required for a patient to improve their symptoms management in this traditional format. Medication and treatment information cannot be reduced to a series of truisms and maintain its accuracy.

A serious video game is one that is designed with a primary purpose other than the players enjoyment, and is most commonly found as a simulation game. It is sometimes very difficult to distinguish between a serious video game and a normal video game. Consider the example of SimCity, a popular city simulation game released by EA in 2013. SimCity can be clearly viewed as a video game yet can potentially be employed as an urban planning tool. Other examples of game which appear to be both serious and not include the popular video game ArmA 2 which is used by the United States Military for combat training and simulation. The serious game Fold-it is possibly the most well known serious game - it allows users to experiment with protein folding in a puzzle solving environment. Players are able to determine the folding pattern of a key protein involved in the replication of the HIV virus, which is included in the video game as a solvable puzzle. Foldit players sometimes outperform state of the art computational algorithms (Khatib, Cooper, Tyka, Xu, Make- den, Popovic, Baker & Players 2011). Serious games feature far more complex game mechanics than an educational game - players are expected to learn to understand the system being modelled by the video game through experimentation.

3 Related Works

This area is not an unexplored one; a large number of educational video game have been developed, including a significant number targeting diabetes education. Projects such as the "Escape from Diab" aim to encourage those who play video games to exercise more and monitor their food consumption more closely, while some other research groups are attempting to introduce gamification into diabetes medication systems (Thompson, Baranowski, Buday, Baranowski, Thompson, Jago & Griffith 2010).

In the videogame the "Escape from Diab", the focus of the project is to encourage consumption of low energy density foodgroups such as fruit and vegetables and to encourage physical activity amongst obese youth. The project attempts to change ingrained behaviours of youth through a process of gradual change in mediators such as skill levels, knowledge and self esteem, which in turn is expected to result in a change in behaviour. The Escape from Diab experiments with a number of different behavioural theories such as social-cognitive theory, self-determination and the Elaboration Likelihood model. These models were used to guide game development decisions and were incorporated where possible with game development guidelines to develop and entertaining and engaging game (Thompson et al. 2010).

Similar projects have been undertaken in the past, such as the historic SNES video game "Captain Novolin" in which the titular character is a superhero who also suffers from type 1 diabetes (Friebberger n.d.). The player must manage the glucose levels of the player character, Captain Novolin and fight enemies which appear visually similar to fast food. The game was sponsored by the Novo Nordisk company - a pharmaceutical group which distribute the Novolin brand of insulin. The effectiveness of the Captian Novolin game at improving diabetes management has not been measured, however investigations into the impact of playing the video game at the time showed that participants felt more comfortable discussing their diabetes with friends and family (DeShazo, Harris & Pratt 2010).

Several projects have taken a different approach and attempted to use gamification in conjunction with blood glucose measurement tools. Didget is one such system (Klingensmith et al. 2011). Consisting of a blood glucose monitor which functions as an attachment to the popular Nintendo DS handheld gaming platform. The Didget device comes with a selection of minigames which require that the player to take their own blood glucose level as a game play element. Additional minigames can be unlocked as a reward for frequently using their glucose monitoring device. An evaluation study carried out in 2011 found that participants who were using the Didget system found it to be motivating and helpful for building good blood monitoring habits (Klingensmith et al. 2011).

Left 4 Dead 2 is not an educational video game. It does, however, include a "Director" - an artificial intelligence capable of redesigning the game world to present new and interesting challenges to the player (Booth 2009). This can take the form of moving item pickups and enemies, to altering the weapons available to the player. Most importantly, it can change the enemies the player must fight to create dramatic tension at key moments and ensure the player is constantly engaged with the game. Such a system could easily be extended to dynamically determine the challenges a player faces based upon their own medical history, allowing a game to be tailored automatically to the player’s particular medical requirements.

4 Advantages of Video Games Over Other Media

Video games as a form of interactive medium have several significant advantages over other forms of diabetes education. One advantage which can be easily seen and potentially will have the most impact is the easy replication of the product. Tens of thousands of diabetes sufferers can be provided with a copy of a video game for no additional cost. In addition to this, it is an known to researchers that providing customized and tailored assistance to patients can significantly improve the quality of their engagement with support programs. Video games can easily be configured to support this customization through dynamic content placement and visual adjustments. For example, a patient with an allergy to seafood could receive a video game with a predetermined configuration which alters the spawn points of food to have less seafood, or to include specific game missions which deal with managing allergies in the contexts of diabetes. This form of dynamic content and configuration is not available in any other medium. Some other ways in which video games are an ideal medium to use when educating patients about Type 1 diabetes relate to the motivation patients feel to play a video game. Video games are incredibly enjoyable and can be made to appeal to a wide range of people, and provide constant positive feedback for patients when they play the video game. It is hoped that this motivation can improve the users engagement with the video game.
5 Feasibility of Video Games as Educational Tools

Video games have a long history as educational tools, dating back to early video games such as Where in the world is Carmen San Diego?. Their effectiveness as educational tools has been established by a number of studies (Rosas, Nussbaum, Cumsille, Mariano, Correa, Flores, Grau, Lagos, López, López et al. 2003) (Brown, Lieberman, Gemeny, Fan, Wilson & Pasta 1997) (Dondlinger 2007). Furthermore, the use of video games for diabetes education has been explored and found to be an effective tool for education about symptom management (DeShazo et al. 2010). We believe additional effectiveness remains untapped and that a more cohesive integration of the educational components of the game and the gameplay elements, similar to the integration of orbital mechanics in Kerbal Space Program and boolean logic in Minecraft's redstone system, will be a more effective method for imparting knowledge to the player. Video games are certainly not a silver bullet and have been unable to influence children in some areas (DeShazo et al. 2010) we believe that video games are a worthwhile and viable education tool. Additionally, commercial video games have an incredible ability to motivate people to play them (Kirriemuir 2002) which provides an extremely effective method for indirectly educating people about in game concepts (Egenfeldt-Nielsen 2006).

6 Video Game Design

This project draws inspiration from the video game Bioshock, in which users are required to manage a resource called EVE. EVE can be used to perform special powers, and is restored by consuming items found in the environment. In order to complete the challenges presented to players in Bioshock, players must effectively manage their EVE levels. By replacing the resource management techniques used to manage EVE with those required to manage blood glucose levels it is hoped that players will learn to use these techniques with a similar level of effectiveness, and then be able to transfer these skills and knowledge to managing real world diabetes symptoms.

Additionally, the player character will experience hyperglycemic and hypoglycemic episodes should their glucose resource exceed or drop below certain thresholds. When this happens, the player character will provide certain cues to the player which indicate that an episode is about to occur. The cues provided to the player will match symptoms which indicate that a hypo/hyperglycemic episode may occur, such as tiredness and lethargy. The player can always see their glucose level relative to the high or low thresholds in the top left of the screen, but it is hoped that by presenting these cues to the player they will begin to associate these feelings and scenarios with the risk of hypo/hyperglycemic episodes and be able to act accordingly.

This style of education is different from that explored by other, similar research and commercial projects. The goal of this project is to provide Type 1 diabetes sufferers with a risk free means to experiment with the management of their symptoms in the hope that the knowledge gained through this experiment can be adapted to real world scenarios. We will achieve this through two mechanisms:

1. Providing a way of experimentation where players are free to consume different forms of food in different quantities, self medicate with insulin and partake in physical activity. It is hoped that players will gain an understanding of how food, insulin and exercise interact to effect blood glucose levels through this unstructured interaction system.

2. Provide a series of scenarios which relate to real world scenarios in which diabetes management may be particularly challenging. It is hoped that by providing these scenarios we will be able to expose players to challenging scenarios and thus equip them to deal with these scenarios in the real world.

As opposed to the explicit inclusion of diabetes content in the projects discussed in the related work section above, this project has a more implicit inclusion of diabetes related content. The goals of the game designers are not readily apparent to the user. We believe that this will encourage more experimentation from users. Additionally, we believe that by integrating blood glucose management as a game mechanic players will be forced to master the skill to progress through the game. This game has been targeted at young adults between 18 and 25 due to the high uptake of video games within that age sector.

7 Blood Glucose Levels as a Resource System

A desired outcome for this video game was to provide players with some experience managing the glucose levels of a character in the hope that these management skills will be transferable to real world scenarios. For this knowledge transfer to be feasible, there must be a correlation between game actions and real world actions. In practice, this means that any and all in-game actions a player can take to manage their blood glucose levels must convey all important information relating to the equivalent real world action. Additionally, the relative effectiveness of different actions must be accurately represented in the video game to prevent one particular action or set of actions becoming a dominant strategy. The presence of a dominant strategy will mean that players do not experiment with all possible actions but instead reuse the same sequence of actions - an undesirable outcome. This process of ensuring the viability of different techniques in video games is colloquially known as “balancing”. With this in mind, we identified a number of actions which form the core of diabetes management techniques based upon current research and medical practices. These actions where then simplified to produce elements which can be easily understood and then introduced into a video game environment. Through several iterations of playtesting, a roughly balanced set of actions where produced. These include physical activity - represented as a slight drain on the players glucose level, and the drain for using special powers - food consumption, which increases the characters glucose levels gradually, and insulin consumption, which can reduce the characters glucose levels.

8 Diabetes Game Mechanics

Based upon user evaluations, current medical practice for diabetes treatment and information about the disease itself (Hanas 2007) we have identified the following set of game mechanics which we believe can be used to form a model of diabetes in a video game.

- A resource representing blood glucose levels
This resource should be increased by consuming food
This resource should be decreased by consuming insulin
This resource should be decreased by performing physical activity.

- A large variety of foods differing in quantity, glycaemic index and carbohydrate content.
- A number of scenarios which challenge the players mastery of this resource system.

Furthermore, we propose that the inclusion of these game mechanics will result in players being more informed about, and having a greater understanding of, their own diabetes symptoms based upon the response to similar systems in other commercial video games.

9 Prototype Development

In order to facilitate rapid development, the video game prototype was produced using a free game engine called Unity 3D. Unity 3D is highly popular and well supported toolkit, and it is hoped that by using an existing engine that later development will be made easier due to the popularity of Unity 3D and the lack of support requirements. The game currently supports a number of features, the most significant of which are listed below.

- The ability of the player and enemies to use and fire weapons. Weapons damage their targets which then die.
- The ability to consume food / other health items to restore health and glucose levels.
- The ability to consume insulin to reduce the players current glucose levels.
- The ability to walk, run, jump, and use special powers which consume glucose. This is the main way in which hypoglycemic episodes are induced.
- Automated pathfinding using an implementation of A*. This is used to assist enemy navigation and movement.
- Simulated Hypo/Hyperglycemic episodes.

- A HUD showing glucose levels and health.

The video game contains a simple level suitable for demonstrating the effectiveness of the game mechanics being employed and as such is populated predominantly with temporary or procedurally generated content. Levels were initially designed with much denser content, however this distracted playtesters from the gameplay mechanics.

A prototype was produced featuring a small amount of gameplay and was used to test users responses to the video game and the mechanics representing the diabetes game mechanics.

10 Prototype Evaluation

During the development of the prototype, a small sample of playtesters were used to examine the viability of the diabetes game mechanics. This viability related to how easy users found it to understand the diabetes game mechanics, how flexible the diabetes game mechanics were, and how much of an impact the diabetes game mechanics has on the gameplay of the prototype. A playtesting session usually lasted for 5-10 minutes and consisted of a user playing the
11 Evaluation of Health Impacts

The purpose of this application is to develop a greater level of understanding in players of how glucose levels are impacted in scenarios represented in the game. It is thus critical to determine how players respond to the application and as such a formal evaluation of the applications impact on diabetes management should be undertaken. An effective evaluation would require that a selection of participants with type 1 diabetes of varying genders and ages play the video game over a small period of time. The abilities of the participants to manage their diabetes symptoms should be measured before and after using the video game. Measurements should relate to the content portrayed in the video game, specifically blood glucose management, calorie counting, portion and meal sizing and exercise levels. These can be measured by investigating the levels of glycaated haemoglobin in test participants, as well as surveys and interviews.

12 Future Work

There exists a clear continuation of this work - the development of a full scale video game production suitable for distribution and testing. There are a number of ways this video game could be developed. The prevalence of mobile devices has resulted in the development of a large mobile gaming market. This market has limited technical capabilities and is not swayed by impressive graphical effects, resulting in a resurgence of 2D games. It is quite possible that a 2D game aimed at mobile devices would be a viable and effective medium for developing a diabetes education video game. However, mobile games require short interaction periods and require create control schemes due to the lack of hardware controls, potentially complicating development. Conversely, it is possible to develop a large scale, high cost and production quality video game aimed at personal computing and console "gamers". This market segment is easier in some ways to develop for as the control schemes are far more standardized and players have longer interaction sessions with the video game, potentially allowing for a greater impact on participants in terms of behavior modification. There is also a possibility of developing a modification for an existing video game. Several video games currently on the market include extensive modification support and tools which would significantly reduce the costs of developing a modification. The cost of developing a modification to an existing game is significantly lower than that of creating a game from scratch, however it requires that users who wish to play the modified game must purchase a copy of the game, limiting distribution. Thus, games which are already commonplace are an ideal target for this practice. Games such as Minecraft and The Elder Scrolls: Skyrim are easy to modify and highly popular, making them potential candidates for modification. A modification to an existing video game would require far less time and resources to develop than a bespoke game, in particular less time requirements for developing game content. Modifications to make these games educational already exist (Short 2012) and have been shown to be a viable tool for creating educational games compared to the development of an entire game from scratch.

The development of a modification to either Minecraft or The Elder Scrolls: Oblivion or The Elder Scrolls: Skyrim is recommended as this would allow for a working product to be obtained with a much lower cost than developing a whole game. For the
purposes of evaluating the effectiveness of the game mechanics identified above, the issues incurred by developing a modification and not a stand alone product are easily overcome, making this practice a viable area for future research into educational video games.

13 Conclusion

We have identified a need to improve symptom management of Type 1 Diabetes and shown that video games can be an effective tool for improving symptom management. We have further identified a set of game mechanics which we believe can serve as a viable model of diabetes in a video game and produced a proof of concept video game to demonstrate these game mechanics. Furthermore, we have identified the development of a modification for the popular video games Minecraft or The Elder Scrolls: Skyrim as potential areas of future research.

14 Acknowledgements

We would like to acknowledge and thank Dr. Ian Warren for his guidance, Dr. Ralph Maddison for his input and Dr. Jim Warren for his encouragement in this project.

15 References

Booth, M. (2009), The ai systems of left 4 dead, in ‘Keynote, Fifth Artificial Intelligence and Interactive Digital Entertainment Conference (AIIDE09)’.


Hanas, R. (2007), Type 1 diabetes in children, adolescents and young adults: how to become an expert on your own diabetes, Class Publishing Ltd.


Social Media and Online Survey: Tools for Knowledge Management in Health Research

MARK MEROLLI, FERNANDO JOSE MARTIN SANCHEZ, KATHLEEN GRAY
Health and Biomedical Informatics Centre
The University of Melbourne
Level 1/202 Berkeley St, Parkville, Victoria 3010
merolli@student.unimelb.edu.au

Abstract

Intro: This paper outlines the design, dissemination and recruitment of participants into a global online survey examining perceptions regarding the value of social media for chronic pain self-management. Building on literature supporting the use of online survey in health research design, we outline the step-by-step process involved in creating this survey and also discuss how use of social media as a distribution channel may significantly impact participant recruitment. Methods: We designed the online survey using Survey Monkey. After the initial survey design phase we consulted a survey expert and statistician to refine the instrument before obtaining ethics and piloting the survey. Recruitment consisted of both traditional online dissemination (i.e. emails and website posting), as well as a more active approach utilizing various social media. Results: 231 people with chronic pain took the survey. After applying exclusion criteria we were left with N = 218 responses for analysis. Demographics and social media platform use are presented, as well as a detailed look at the survey recruitment process, contrasting traditional online recruitment with that using social media. Conclusion: Online surveys are a valuable study method for health research. They build on the foundations of traditional survey method but harness the power of the Internet to conduct research. Concurrently, social media provide a novel distribution channel for survey recruitment and their potential application in health research is worthy of further consideration.

Keywords: Online Survey, Social Media, Information-Communication Technology (ICT), Study Recruitment, Therapeutic Affordances, Chronic Disease

1 Introduction

We recently conducted a global online survey to obtain a more detailed understanding of how social media may influence health outcomes in chronic pain self-management and also, to garner individual perceptions regarding social media’s underlying therapeutic value. The survey built upon findings from literature review we conducted that was published elsewhere, investigating the health outcomes and related effects of social media use in chronic disease (Merolli, Gray, Martin-Sanchez 2013). In this same review, we were able to qualitatively identify and categorise various perceived therapeutic affordances of social media that may underpin their value to chronic disease self-management. We have labelled these: Identity, flexibility, structure, narration and adaptation. Perceived affordance theory dictates that the user perceives the potential for action of how an object may be used, thus predicting and influencing how the ensuing interaction proceeds. In more recent years, information systems researchers have begun to see the potential application affordance theory may have to the study of human-computer interaction (HCI). Research continues to be published investigating how perceived affordance may be a valid model informing information and communication technology (ICT) design (Sutcliffe, Gonzalez & Binder et al. 2011, Anderson 2011, Zhao, Liu & Tang et al. 2013).

The survey presented in this paper aims to refine our understanding of therapeutic affordances so that we may look to clinically test their influence on improving health outcomes in chronic disease self-management. Studying chronic pain, this paper outlines the online survey instrument design, piloting and recruitment process, while also presenting a case for the power of social media for study recruitment purposes in health research.

1.1 Online survey in health research

Online surveys are relatively common for the study of Internet usage in healthcare. Examples already exist of studies employing online survey to examine Internet and social media usage patterns in chronic disease management (van Uden-Kraan, Drossaert & Taal et al. 2009, Corcoran, Haigh & Seabrook et al. 2010, Bartlett & Coulson 2011). Online survey has many advantages. Some of which include: the ability to be selective of participants, the ability to allow public access and participation, they allow participants to remain anonymous and they are convenient for the research team.
in allowing responses to be stored in a readily accessible database (Eysenbach & Wyatt 2002, Daley, McDermott, McCormack Brown et al. 2003). Eysenbach and Wyatt (2002) report that many studies have been published supporting the validity and reliability of online survey by contrasting and correlating results to traditional offline approaches.

Surveying participants allows for sufficient comparison of observations to be conducted in health research. In the case of the present research, it allows for identification of trends in social media usage (i.e. how social media usage may change health outcomes). It also allows correlation to be made investigating certain relationships (such as, analysis of whether effects/outcomes are linked to the therapeutic affordances of social media) (Saks & Allsop 2007).

However, whilst online surveying may help demonstrate a link between social media, its therapeutic affordances, and health outcomes, it does not in itself provide enough evidence to insinuate a relationship. For a survey to be successful in providing meaningful descriptive data it is crucial the right questions are formulated and scrutinized. In some ways, the researcher should have a good idea of the answers to expect before beginning (Gable 1994).

We take this opportunity to outline a novel research method for online survey design in health research and describe how the use of social media can influence the participant recruitment process.

2 Methods

Given that no matching survey design exists that has been validated in chronic disease management, our survey presents a novel approach and design. However, design components were based on best available evidence and attempted to build on the foundations of other online surveys within this domain. The survey serves two main purposes. Firstly, to refine our understanding and examine the presence of the aforementioned therapeutic affordances and their value to health intervention design and secondly, examine their potential to underpin improved health outcomes clinically via online interventions in future research.

2.1 The Instrument

The survey was developed and hosted on survey platform, ‘Survey Monkey’ (Survey Monkey is an open-access survey creation and hosting platform). It consisted of a self-administered online survey of 240 questions, obtaining participant demographics, their chronic pain and disease relevant information, current health symptom status, social media use, perceptions towards the value of various therapeutic affordances and reported effects/outcomes from social media use. There are other open-source survey platforms available to choose from and one is not limited to Survey Monkey. Another popular platform considered was ‘Lime Survey’. It enables many of the same survey functionalities and processes as Survey Monkey, such as complex question logic, in-built analysis capability and the ability to produce visuals of survey results. Both possessed appropriate functionalities to meet the complexities of our survey design, clean and user-friendly interfaces, charts/graphs, and also ability to export data to programs such as SPSS statistics software for more formal analysis. Ultimately, Survey Monkey was chosen due to the previous experiences of the study authors and comfort with the platform. This may differ for other researchers.

2.2 Survey Design

The online survey (available on request from the corresponding author) is broken down into four major areas (demographics, chronic pain/disease information, patient-reported health status, social media use/therapeutic affordances perception and outcomes from use. Given the complex nature of the survey and a desire to minimize survey fatigue, we used various question logic and skip options enabled by Survey Monkey to guide respondents through the survey and only elicit necessary information. We used Likert rating scales ranging from “strongly agree” to “disagree” and “not at all” to “very often” as per previous validated survey models (WHO 2002, Cella, Riley & Stone et al. 2010, Fox 2010, Fox 2011).

When ethics approval was sought, we made it clear that the survey aimed to obtain anonymous data pertaining to social media use for chronic pain management to better protect participant’s privacy. In order to respect the participant’s right to participate free of coercion we also made it explicitly clear that participation was voluntary, the survey did not offer medical advice, nor did it insinuate a medical relationship between patient and researcher. Before final ethics clearance was given, we were asked by the Human Research Ethics Office to avoid use of the term ‘affordance’ in the study title as not to confuse participants and also to copyedit the Plain Language Statement (PLS) to suit online viewing.

Before opening the survey, we sought consultation with a survey design expert and statistical consultant from the Statistical Consulting Centre (SCC) at the University of Melbourne to better understand the intricacies of survey design. Over three meetings discussions centred on sample size and statistical power, survey length, question style, eliminating bias and finally recruitment methods. Major technical and design issues to come out of these meetings included: consistency of statement/question structure, using the same number of statements for each platform, statistical power calculations, taking care to avoid use of double negatives as well as leading statements and finally, employing question skip logic to decrease survey fatigue.

Addressing survey fatigue was paramount. Our statistical consultant commented that 15-20 mins was appropriate survey length. However, he noted that on average survey participants lose focus the further they move into a survey, thus, suggested that we place the social media platforms of most interest to our research earlier in the survey. We were more interested in investigating social media platforms compared to earlier social technologies as contrasted and discussed in Merolli
et al. (2013), therefore platforms such as social network sites (SNS), blogs and wikis were placed earlier in the survey as opposed to discussion forums and chat rooms.

Statistical power and dissemination were another major focus of these meetings. Social media use to improve health outcomes in chronic disease management is still a relatively uncharted area. Previous research and informed research designs are still in their infancy. For this reason, calculating desired sample size was problematic. We referred to other online surveys created within a similar research context and there has been no specific validated sample size in any of these studies surveying social media usage in a chronic disease setting. Study methods employed varied considerably and nor was there any mention of how these studies justified their own sample sizes (Chung & Kim 2008, van Uden-Kraan et al. 2008, van Uden-Kraan et al. 2009, Bartlett & Coulson 2011, Setoyama, Yamazaki & Namayama 2011, Klemm 2012, Mo & Coulson 2012). Sample sizes ranged from N = 32 in van Uden-Kraan et al. (2008), through to N = 528 in van Uden-Kraan et al. (2009). The mean of all studies being N = 255. For the above reasons, we aimed to achieve a final sample around this range, aiming for 200-250 responses. Finally, we were warned of the inherently poor response rates to surveys. Although online surveys allow for a greater potential reach and spread of responses, data collection methods would require substantial consideration to achieve power.

2.3 Survey Piloting

Between our second and final meeting with a survey and statistics consultant, we planned a pilot of the survey in order to address any usability and design issues. Once ethics clearance was received we sought input from both social media using patients and technology experts from the Department of Computing and Information Systems at the University of Melbourne. Piloting was conducted to assess survey design and whether the quantity and quality of questions/statements was satisfactory. Usability was assessed inline with the adapted work of Bevan (2009), who discussed definitions of usability in information communication technology (ICT) design and postulated a framework for its measurement in information technology applications. The assessment of usability surrounds measurement of: Attractiveness & aesthetics, suitability & appropriateness of functions, ease of use & user interface design, learnability, technical issues and safety & security of design. We had a total of five patients, and three computing and information system’s staff pilot the survey. A questionnaire was embedded into the end of the beta version of the survey (a summary of questions is available on request from the corresponding author).

2.3.1 The Survey Domains

a) Participant Demographics (Q. 1-8). The first survey domain asked participants general demographic questions (e.g. gender, age, education, employment, etc.) and were adapted from the World Health Organization’s ‘World Health Survey’ (WHO 2002).

b) Chronic Pain/Disease Information (Q. 9-13). The second domain of the survey asked about current chronic pain/disease status. Examples were: “Do you suffer from chronic pain (pain over 3 months duration)?” and “Have you been formally diagnosed with a chronic disease that has led to your pain? and ‘If yes’, what is the condition?”, etc.

c) Health Status (Q. 14-30). We used items from the “pain interference” item bank of the Patient Reported Outcome Measurement Information System (PROMIS) to measure Health-Related Quality of Life (HRQL) (described in Amtmann, Cook & Jensen et al. 2010). We also included one “pain behaviour” item that measured pain severity via a visual analogue rating scale. This is common amongst chronic pain studies for the measurement of pain intensity (Dworkin, Turk & Farrar et al. 2005). Selecting a Validated Patient-Reported Outcome (PRO) Measurement Tool was a crucial requisite. PROMIS - Pain Interference (PROMIS-P1) was chosen to overcome consistency issues with pre-existing pain outcome measures. Common legacy measures can be unyielding, requiring respondents to complete every item even when items provide little to no extra useful information about pain interference (Amtmann et al. 2010). PROMIS provides ‘item-banks’ to measure outcomes of interest. Using an ‘item response theory (IRT)’ model, direct comparison of scores can be made even when different items are selected, thus allowing for better across-time and different sample comparisons (Cella et al. 2010). However, perhaps most significant, allows for the development of comparable and flexible short-forms specific to a target population or study (Amtmann, et al. 2010, Gershon 2012, Witter 2012). Custom short forms can be developed for specific purposes and/or samples (Amtmann et al. 2010). We selected items from the PROMIS-P1 to represent all major sub domains of pain interference: Cognitive affect, sleep, recreation/leisure activities, social life, activities of daily living (ADLs), psychosocial health and physical health.

d) Social Media Use by Chronic Pain Sufferers (Q. 31-240). Given the survey’s focus on understanding social media use for health self-management, this was the most comprehensive, time and labour intensive section of the survey. Participants were asked what social media they had used as part of chronic pain self-management (e.g. “In the last year, have you used Social Network Sites when you go online for information, communication or interaction about your chronic pain?”). We also asked about frequency of use, the features of the platform they used and self-reported effect use has had on their condition. The remaining questions pertained to the therapeutic affordances of social media that we put forward. Each affordance consisted of three statements designed to observe the strength and degree to which the platform in question perpetuated that affordance and nurtured changes in health status. All social media related questions were the same across all platforms to ensure reliability and consistency. As previously eluded to, given the ambiguous nature of affordances outside of academic discourse, statements designed to measure the value of therapeutic affordances were phrased in such a way as not to confuse or bias participants and therefore made no
mention of the word ‘affordance’, instead using tags such as ‘value’ and ‘prefer’.

2.4 Recruitment/Data Collection

A Google search was performed periodically from March 1st through to May 20th, 2013 to identify potential distribution channels. Search was limited to groups and organisations in the English language, with chronic disease or chronic pain the focus. We targeted large online health networks due to their high user bases (e.g. Patients Like Me, Daily Strength), smaller online pain support communities and global chronic disease and pain organizations (e.g. Chronic Pain Australia, Pain Research Forum UK, American Cancer Society, etc.). In addition we later searched common health online social networks and targeted active chronic pain related groups on Facebook and Twitter. Recruitment was augmented by contacting various other influencers in the field, as well as posting to active social media in health research groups on LinkedIn.

When we emailed each organization or group moderator it was made clear that our survey focussed on “pain interference” as a result of living with the condition his or her group supported (e.g. arthritis, diabetes, fibromyalgia, cancer, etc.). Formal invitation emails were sent to moderators requesting they pass study details onto members to participate in the survey. A recruitment video was also created on ‘Animoto’ to describing the study and the link was pasted into the email text (Merolli 2013). We did not offer financial incentive to participate. However, we did offer to provide preliminary survey results in appreciation of support. This is advocated by Eysenbach and Wyatt (2002), who explain that despite contributing to selection bias, is less likely to seriously skew results compared to cash incentives. The PLS and informed consent pages were embedded into the survey. By doing this, moderators were able to view the suitability of the survey to their members before agreeing to support the work. We asked that if organizations/moderators were willing to disseminate the survey amongst members and/or clinical and research colleagues, they post the survey on their websites and/or to social media accounts where appropriate. The survey was open until July 1st, 2013. If participants wished to participate, the survey link led them to the PLS, which outlined inclusion/exclusion criteria (age of eighteen years or over, have chronic pain and have used social media as part of chronic pain self-management). This in-turn led to the informed consent page, which then made the actual survey available once participants clicked to agree to terms.

3 Results

We collected data for N = 231 individuals, of which N = 4 did not complete the mandatory first question regarding living with chronic pain. Therefore, N = 227 people answered. N = 9 did not meet eligibility, selecting “no” to suffering chronic pain (automatic survey logic excluded these people), leaving a total of N = 218. Summary statistics for demographics of survey participants can be seen (see Table 1). Participants were predominately female. Age ranged from forty to forty-nine. Many were married and at least high school educated. However, a large number were not working for pay due to ill health. Most indicated that they had been diagnosed with a chronic disease leading to their pain (most reported condition was fibromyalgia). Social media platforms used in self-management are listed (see Figure 1). SNS were the stand out, followed by discussion forums and blogs.

The results of this survey also highlight the wide reaching potential of the Internet to disseminate an online survey. The global spread of our results was testament to the power of the Internet to recruit from around the globe and give a truly diverse data set (see Figure 2). We were not surprised to see that the majority of respondents came from Australia given the survey’s origins (N = 128). However, we also received a respectable number of responses from the United States (N = 41) and United Kingdom (N = 23). Followed by Spain (N = 8), Canada (N = 5) and New Zealand (N = 4). Small numbers of responses (i.e. 1-2) also filtered through from Ireland, South Africa, China, Kenya, Pakistan, Burma and Taiwan.

Table 1: Participant Demographics

<table>
<thead>
<tr>
<th>Gender</th>
<th>% of Responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>16</td>
</tr>
<tr>
<td>Female</td>
<td>84</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Age Range</th>
<th>% of Responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>18-29</td>
<td>17</td>
</tr>
<tr>
<td>30-39</td>
<td>22</td>
</tr>
<tr>
<td>40-49</td>
<td>31</td>
</tr>
<tr>
<td>50-59</td>
<td>21</td>
</tr>
<tr>
<td>60+</td>
<td>9</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Marital Status</th>
<th>% of Responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Never Married</td>
<td>22</td>
</tr>
<tr>
<td>Currently Married/Partnered</td>
<td>59</td>
</tr>
<tr>
<td>Separated/Divorced/Widowed</td>
<td>19</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Level of Education</th>
<th>% of Responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>High School or Less</td>
<td>33</td>
</tr>
<tr>
<td>College/University Completed</td>
<td>42</td>
</tr>
<tr>
<td>Post-Graduate Degree Completed</td>
<td>25</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>If Not Working For Pay (Reason?)</th>
<th>% of Responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ill Health</td>
<td>76</td>
</tr>
<tr>
<td>Other</td>
<td>24</td>
</tr>
</tbody>
</table>

Table 1: Participant Demographics
3.1 Examining Modes of Dissemination

The survey was opened on May 24th. At this time emails were sent to various chronic disease and pain organizations asking for support to disseminate the survey to group members. We posted a recruitment blog via our departmental website (HaBIC 2013), posted the study to our Facebook page and Twitter account, as well as to Scoop.it. Figure 3 represents the progress of survey responses collected over time. From what can be seen in the figure, response to our survey was initially quite slow. We calculated that in the first week of recruitment, we received only N = 29 responses. Week 1-2 represent a focus on more traditional means of online recruitment whereby no active sharing of the study was taking place. For example, we posted the study to organisations/moderators, posted it to our blog and posted to our Facebook page. All of which give initial limited visibility but fail to a) hone in on our target audience (patients) and, b) generate interest enough to share the survey. From the beginning of week 2 through to early week 3 we saw a steady increase in the number of weekly responses, more than doubling and peaking at N = 72. At this time we were able to observe several factors that we believe contributed to the increased response rate: a) the survey started to appear on various chronic disease/pain...
website news feeds, Facebook pages and Twitter accounts and, b) we altered our recruitment effort by contacting online support groups on Facebook directly, rather than continuing to rely heavily on email and, c) we began to receive emails, notifications and tweets from people who had shared the survey via their own social networks and other platforms. Further progress from week 3 through to week 5 (to when the survey closed on July 1st) showed a steady stream of responses to continue to come in with week 4 bringing N = 54 responses, before one final peak of N = 66 just before close. In the final week of recruitment, we made one last attempt to recruit participants by posting to social media in healthcare special groups on LinkedIn.

![Survey Responses vs Time](image)

Figure 3: Survey Responses vs. Time

4 Discussion

4.1 Considerations For Conducting Online Surveys in Health Research

Eysenbach and Wyatt (2002) provide instances where online survey may be highly suitable in health research. For instance, when target participants are already enthusiastic Internet users (as was the case in our own inclusion criteria) and when a global spread of responses is desired (see Figure 2). Online survey is also advocated in survey design to keep costs down and also when the survey is complex in nature, requiring branching and question logic to guide respondents through (Daley et al. 2003). This same complexity was present in the present survey and as previously discussed supports our choice to use Survey Monkey. Eysenbach and Wyatt (2002) also state that the power of online survey is recognised when large amounts of data need to be collected and analysed in a timely fashion. This is particularly relevant to our study, with N = 231 responses to 240 questions. We were able to feed results directly into statistics software, SPSS.

However, Eysenbach and Wyatt (2002) highlight a few particularly salient precautions when considering the appropriateness of online surveys for health research. Online surveys are well recognised in qualitative research. However, for quantitative purposes researchers need to take into consideration potential bias that comes with the online survey method. Often the target group (i.e. chronic disease sufferers) is underrepresented on the Internet, thus leading to selection bias (Eysenbach & Wyatt 2002, Daley et al. 2003). This fact has previously been described in chronic disease and Internet research published elsewhere (Fox 2010). Self-selection bias is also discussed, whereby participants are more likely to respond to a survey when the study area is of direct interest, potentially creating a skewed sample.

4.2 Opportunities For Social Media In Survey Recruitment

Results presented showing progress of our survey recruitment highlight two similar, yet contrasting modes of online participant recruitment. Week 1 of survey recruitment represents a traditional (Web 1.0) online recruitment style, where despite posting to a variety of media, visibility was inherently low with minimal spread or call for people to be actively involved (Close, Smaldone & Fennoy et al. 2013). This mode of recruitment included: Email, blog posts and the survey website itself on Survey Monkey. From week 2 and beyond, recruitment style shifted and relied much more heavily on the power of social media to generate activity and sharing, as well as active participation by the public in the recruitment process. Close et al. (2013) suggest that this is because along with greater numbers using social media comes increased visibility and potential for communication. We recognised a much more consistent inflow of survey responses at this point as the survey worked its way across the social web and essentially became part of the digital conversation, rather than laying...
in waiting for people to stumble across it. We saw one such instance, where the survey was posted to the Facebook page of a chronic pain clinical research team. It was commented on, liked and shared by a handful of members to their own pages, some of which we then further shared by other people in extended networks. All this occurred over a matter of hours.

Close et al. (2013) introduce the term ‘snowballing’ to describe this type of study recruitment via social media, as we have described above. This highlights the spread of study information across online social networks, thus increasing reach, visibility and ultimately participant numbers. Furthermore, Close et al. also suggests that increased visibility across multiple social media augments traditional online recruitment advertisements. In our case, advertising of the survey on our departmental website, the survey website itself and the emails that were sent, were augmented by blogging, Facebook, Twitter and LinkedIn efforts. This represents a layering of the recruitment message and ultimately has each medium supporting the other and, optimizes the likelihood that recruitment will be successful. Layering also has potential to improve trust between researchers, the research institution and potential participants. In providing multiple channels for exposure, our supporters and participants were given the opportunity to digest and ponder participation in our survey in more depth, creating sense of autonomy and ownership over their eventual participation.

On the other hand, Close et al. (2013) also indicate that care must be taken when using social media for study recruitment purposes because often the presence of clinical researchers may be a sensitive issue to online patient communities gathered around a specific disease. If online community members sense the call for research participation does not reflect the goals of the community, mistrust develops and forms a barrier to recruitment and participation. In order to avoid this and limit any potential negative affects on these patient communities we deliberately did not contact individual patients directly, nor did we unscrupulously and pervasively post recruitment messages in any communities without invitation. All study advertising on social accounts that were not our own was done by the disease organisations’ own staff or group moderators.

Finally, results from Fenner, Garland & Moore et al. (2012) show strong predictors of participation in online studies are dictated by various participant demographics. These include: Older age, being female, a higher level of education and coming from a higher socioeconomic status group. We too have found similar demographic profiles of online health seeking people and this description of the more typical health seeking social media user has previously been reported (Fox 2010, Meroli et al 2013). Particular studies have focussed their efforts on the effect of various demographics on study participation, particularly as related to gender (Mo, Malik & Coulson 2009). Despite being beyond the scope of the present study, it is worthy to highlight the potential impact of gender on study recruitment and survey results. Although we were not able to verify the male/female membership ratio of the participating organisations, results (see Table 1) indicate that females were the more active survey participants, thus creating a skew of survey findings to the female population. We are currently investigating whether gender played a role in the patient-reported health outcomes obtained in the survey, which we hope to present at a later date. However, we do believe that the present study is reasonably representative of the current online health seeking landscape as described in Mo et al. (2009). Mo et al. present inconsistencies among reports of gender difference in online health communication but one suggestion is that regardless of gender, online patient communities may encourage a greater number of individuals to participate in health self-management with less stigma attached. However, salient to our study, when gender is considered females appear to be more likely to seek health information online and be active participants health research.

Finally, in a similar fashion to Eysenbach and Wyatt’s (2002) description of online survey, Close et al. (2013) report that a pivotal limitation of online recruitment in general is the bias negatively effecting socioeconomically disadvantaged groups and culturally diverse minorities with lack of Internet access. This again raises the question of access, whereby those who may have the most to benefit from social media may be under-represented in studies.

5 Conclusion

This study has highlighted the potential for application of online survey in health research and presents several design considerations. It has also depicted a case that outlines and endorses the possible value of social media to study recruitment. Results, while partly anecdotal, suggest that the wide reach of social media for communication may be ideal for study recruitment. This needs further investigation and may provide a means to deal with the inherent poor response rate that surveys suffer from. Further research is warranted.

6 References


J Med Internet Res 13(4).


BiomRKRS: A Biomarker Retrieval and Knowledge Reasoning System

BAHADORREZA OFOGHI 1,2, GUILLERMO HUGO LOPEZ CAMPOS 2, KARIN VERSPOOR 1,3, FERNANDO JOSE MARTIN SANCHEZ 2

1 National ICT Australia, Victoria Research Lab
2 Health & Biomedical Informatics Centre
3 Department of Computing and Information Systems
The University of Melbourne
Victoria 3010, Australia
{bahadorreza.ofoghi,karin.verspoor}@nicta.com.au
{guillermo.lopez,fjms}@unimelb.edu.au

Abstract
The need for a system to effectively manage and retrieve biomarker information has become apparent to medical and biomedical scientists, as evidenced by the recent development of a number of biomarker information systems. To improve the functionality of such systems, we have developed a new biomarker information system will be discussed in this paper, a system that we refer to as BiomRKRS: A Biomarker Retrieval and Knowledge Reasoning System. In this paper, we introduce the general structure and characteristics of BiomRKRS. We will demonstrate how BiomRKRS employs existing ontologies in the biomedical domain to create a core integrated ontology for biomarkers as a standard vocabulary set for data storage and retrieval. When fully implemented, BiomRKRS will have functionality and utility that will far exceed that of related existing systems due to the incorporation of a knowledge reasoning system that will make logical and useful inferences in the process of semantically processing end-user queries.

Keywords: Biomarker, Ontology, Data Retrieval

1 Introduction
Biomarkers have become central to the current practice of medicine and are an active focus for biomedical and translational research (Olson, Robinson, & Giffin, 2009). The term biomarker (biological marker) was first introduced as a Medical Subject Heading (MeSH) term in 1989 as measurable and quantifiable biological parameters (e.g., specific enzyme concentration, specific hormone concentration, specific gene phenotype distribution in a population, presence of biological substances) which serve as indices for health- and physiology-related assessments, such as disease risk, psychiatric disorders, environmental exposure and its effects, disease diagnosis, metabolic processes, substance abuse, pregnancy, cell line development, epidemiologic studies, etc. (Vasan, 2006). The US National Institutes of Health defines a biomarker as a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacological responses to a therapeutic intervention (BDW, 2001). We specifically adopt the latter definition for its brevity and conciseness. Focusing on Alzheimer's Disease and Multiple Sclerosis, Younesi et al. (2012) proposed that biomarkers may represent molecular, physiological, or structural features and therefore, can be in the form of genes, proteins, DNA, RNA, genetic changes (e.g., SNPs), blood cholesterol levels, or patterns of brain abnormality. Going beyond this definition, our focus is on biomarkers generally in molecular entities including proteins, DNA, RNA, metabolites, and all of the subclassifications of these categories.

Biomarkers have been used for diagnosis, treatment, prognosis, and staging of different categories of diseases, examples of which include the biomarkers for management of postmenopausal osteoporosis (Szulc & Delmas, 2008), prediction and monitoring of osteoporosis (Vasikaran et al., 2011), diagnosis and prognosis of rheumatoid arthritis (Carrasco & Barton, 2010), prognosis and prediction of breast cancer (Weigel & Dowsett, 2010), and treatment of cardiovascular disease (Vasan, 2006), just to name a few.

In addition to directly disease-related procedures, another particularly valuable use of biomarkers is in bridging the gap between the preclinical and clinical development of drugs and vaccines (Olson et al., 2009). Biomarkers can play a role in toxicity/adverse reaction prediction and the analysis of the therapeutic effectiveness of drugs, e.g., dose-response relationship analysis.
In recent years, the U.S. Department of Energy Human Genome Project\(^2\) and advances in genomic sequencing have enabled the detection of new biological features and entities that have been proposed as potential biomarkers for disease diagnosis, treatment, prognosis, and staging as well as for drug development. As a result, the size of the information being generated is increasing every year. To briefly demonstrate this growth of the amount of information related to biomarkers, we have extracted the total number of biomarker-related articles in PubMed from 1947 to 2012 using a simple keyword query. Figure 1 shows the result of this analysis in which the cumulative number of biomarker-related articles totals 638,885 for the specified time period. The diagram also shows how the number of these articles has substantially increased in recent years.

The growth in the number of studies addressing biomarkers necessitates the existence of highly efficient and effective information systems that enable fast and accurate search of and access to up-to-date biomarker information. Our definition of a biomarker information system is a system that stores actual instances of biomarker information/data records as related to different contextual information attributes, e.g., the disease, clinical purpose, and molecular entity, just to name a few. Such a system differs from databases of patient-related clinical information records in that the proposed biomarker information system does not include any patient-specific data but rather captures the background knowledge that relates to known biomarkers. This information can be applied to interpret patient-specific data, but it is itself at a higher level of abstraction.

Some researchers have already started to develop biomarker databases and/or information systems for specific (categories of) diseases and drugs, including the following: a commercial collection of clinical, preclinical, and exploratory biomarkers named GVK BIO Online Biomarker Database (GOBIOM) (GVK Biosciences, 2013), the collection of validated molecular biomarkers in BiomarkerBase (Amlion Research, 2013), the set of population specific and clinically important biomarkers in Biomarker Databases (Liatris Biosciences LLP, 2013), the collection of biomarkers combined with related drugs, targets and genes in the evolvus Biomarkers Database (Evolvus, 2013), the standardized terminology and classification of biomarkers into lifecycle phases and disciplines in the biomarkers module of Thomson Reuters Integrity (Thomson Reuters, 2013), the set of diagnostic and prognostic cancer biomarkers extracted from patents, research articles, and meeting abstracts in the SciClips’ Cancer Biomarker Database (SciClips, 2013), the knowledge-based interface for biomarkers for diagnosis, detection, protection, and characterization of infectious diseases developed in the Infectious Disease Biomarker Database (Yang et al., 2008), the services architecture for the capture, processing, management, and distribution of information in biomarker discovery and validation developed by Crichton et al. (2006), and the ontology representing concepts related to imaging biomarkers developed as Quantitative Imaging Biomarker Ontology (Buckler et al., 2013).

We found that the two major shortcomings of all the above information systems or knowledge-bases for biomarker data management are the following:

- The existing information systems make limited use of standard controlled vocabularies and ontologies for a comprehensive set of features that relate to biomarkers.
- None of the existing information systems makes use of logical reasoning functionalities available in the semantic web domain to semantically process user queries and retrieve (related) information for biomarker-focused user queries.

The advantage of adopting standard terminology and ontologies is two fold: i) the end-users (i.e., clinicians and biomedical researchers) will be able to interact with the system in a controlled manner where the concepts are well-established and well-known to the users, and ii) the system itself will have the ability to make use of reasoning engines over the well-formed structures of existing knowledge-bases. This in turn will enable the system to make inferences and retrieve not only exact matches to user queries but also logically related biomarker data. The latter can play an important role in identification of new indications to previously unknown biomarkers under certain conditions.

\(^2\) http://www.ornl.gov/hgmis

**Figure 1.** The cumulative number of articles published in PubMed in relation to biomarkers since 1947. The PubMed query is: "biological markers"[MeSH Terms] OR "biological"[All Fields] AND ("markers"[All Fields] OR "marker"[All Fields]) OR "biological markers"[All Fields] OR "biological marker"[All Fields] OR "biomarker"[All Fields] OR "biomarkers"[All Fields] AND ('0001/01/01'[PDAT] : "2012/12/31"[PDAT])
In developing BiomRKRS (pronounced similar to biomarkers), consideration has been given to those two main drawbacks of existing biomarker information systems. In the next sections, the general structure of BiomRKRS will be introduced first and then, the main modules of the system will be discussed in more detail.

2 BiomRKRS: architecture

In order to address the two main shortcomings of existing biomarker information systems, BiomRKRS has been developed with a general architecture as shown in Figure 2. The system defines a core integrated biomarker ontology and expanded each core concept through reuse of a number of external ontologies. The core concepts and the external resources that relate to them are described in the next section.

There are some other internal concepts and related logic that are also created in order to construct the core ontology. This core ontology is then used mainly as: i) a (controlled) vocabulary resource for data storage and retrieval, and ii) a knowledge resource for inference purposes in BiomRKRS to semantically process end user queries.

The system implements several functionalities including a knowledge reasoner, a query processor, a data manager, a user manager, and a user interface. These components interact with each other as well as with the core ontology and the system database to answer information requests submitted by the end user. The system database stores three basic types of data: semantic data related to the core ontology, transactional data regarding system users and history, biomarker instance data, i.e., specific information about individual biomarkers.

In the following sections, the main components of BiomRKRS will be discussed in more detail.

2.1 The BiomRKRS biomarker ontology

Based on our expert knowledge, a number of concepts have been defined as related to the main and focal biomarker concept in the core ontology we have built for BiomRKRS. For each of these concepts, a specific external resource, i.e., an ontology in most cases, has been identified to provide specific terminology for the concept. Figure 3 illustrates the core ontology in BiomRKRS including the main concepts, their relationships with the focal biomarker concept, and the external ontologies and knowledge resources utilized for each core concept. In most cases, the external ontology or knowledge resource has been imported into the core ontology and then necessary relations have been created using OWL/XML statements between the external resource/ontology and the specific concept in the core ontology (indicated with an imp in Figure 3 on the links between an external resource and a main concept). In some other cases, the external ontology or resource has been used only as a reference and the actual concepts have been internally created in our ontology (indicated with a ref in Figure 3). More details on how the BiomRKRS ontology is constructed will be given in the next section. At this stage, the main concepts of the BiomRKRS core ontology include:

disease: the disease for which the biomarker is used for diagnosis or prognosis. For this concept, the International Classification of Diseases ICD-10 (World Health Organization, 1992) is used as the standard vocabulary.

endpoint: indicates whether the biomarker is a clinical or surrogate endpoint.

molecular entity: the main entity that is clinically measured as the main biomarker. For this concept, the lists of recommendations for molecular entities from the HGVS (Human Genome Variation Society, 2013) and HGN terminology from HUGO Gene Nomenclature...
Committee at the European Bioinformatics Institute (Gray et al., 2013) are used as the reference set, especially for the DNA and RNA entities.

**pathway:** the biological and genetic mechanism related to the specific disease for which the biomarker is measured.

**purpose:** the clinical purpose of measuring the specific biomarker. For a list of existing purposes, the Quantitative Imaging Biomarker Ontology (Buckler et al., 2013) is used. This ontology includes diagnosis, disease staging, and prognosis as purposes, *inter alia*.

**target:** the sample from the patient to be used in the clinical trial to measure the biomarker. For this concept, SNOMED Clinical Terms (Cornet & de Deizier, 2008) under the "specimen" term are used as the vocabulary.

**technique:** which represents the clinical technique used for measuring the specific biomarker in the specific target. The "method types" from Logical Observation Identifiers Names and Codes (LOINC) (Forrey et al., 1996) are used to expand this BiomRKRS concept.

**validation status:** the stage of validation and qualification of a biomarker. This attribute can have a value of *biologically validated*, *clinically validated*, *in research*, *proposed*, and *qualified* stages.

**stratification criteria:** characteristics that that could affect the validity or measure of the specific biomarker in patients. This includes the age group, race, and gender to which the specific biomarker is related as well as the environmental exposure factors. Concepts under the population term from the Experimental Factor Ontology (Malone et al., 2010) are used for the *race* concept in BiomRKRS. For environmental exposure, related concepts from the International Classification for Nursing Practice Ontology (International Council Of Nurses, 2013) are used.

**evidence:** the source of the biomarker information, e.g. related literature that suggests whether the specific biomarker is, with the specified validation status, to be considered as a measure or not to be used as a measure in the given context for the specific disease. There are also evidence codes defined in BiomRKRS similar to those from the Gene Ontology (The Gene Ontology Consortium, 2000), including computational evidence, experimental evidence, and their subclasses.

### 2.2 Ontology management system

Creating the BiomRKRS core biomarker ontology is the main task carried out by the ontology management system. This involves three main steps:

- Defining and constructing the core ontology, its main concepts, and necessary relationships, discussed in previous section.
- Defining and constructing the internal sub-ontologies for each core concept based on expert knowledge or reference resources.
- Importing external ontologies or knowledge resources and relating them to specific core concepts.

All the above tasks are performed in an automated function to make updates possible as new versions of external resources become available. End users will be able initiate updates through the user interface component, with appropriate access levels and permissions.

The core ontology and the main concepts are created using Web Ontology Language (OWL) in the OWL/XML format. For integration purposes, the external resources...
have also been imported in the same format. In cases where an OWL format was not readily available, we developed a converter functionality in (or prior to) the ontology management system.

In particular, we developed a lightweight converter function for the ICD-10 database to convert the database from its original format, i.e., Classification Markup Language (CiML), into OWL. We made use of instructions given in (Moller, Sonntag, & Ernst, 2010) for most of the conversion procedure. We also developed a converter function to convert LOINC's Tab Delimited file into a Resource Description Format file as part of the ontology management system.

2.3 System components

As shown in Figure 2, there are five internal functionalities in BiomRKRS, implemented in separate components. These components handle accessing the BiomRKRS biomarker ontology, the system database, and interacting with end users.

Knowledge reasoner implements Resource Description Framework Schema (RDFS) reasoning capabilities. For this, the knowledge reasoner has access to the BiomRKRS ontology and its core and integrated concepts.

Query processor is the component that takes in original user query keywords and generates an expanded query. The expanded query includes the original terms of the user query as well as all keywords that are found related to the query term by using the knowledge reasoner component. Figure 4 shows an example query term expansion using the rdfs:subClassOf inference mechanism in RDFS. The selected disease category from the ICD-10 terms, i.e., Benign lipomatous neoplasm of intra-abdominal organs [D17.5], has been found to be related to the three upper ICD-10 disease categories as well as the parent core concept "disease" from the BiomRKRS core ontology. The query processor component creates the expanded query using the "OR" operator between all the disease categories.

Data manager is the component that has been implemented to interact with the user interface and the system database. The data that this component handles are related to the system's history, ontology files and repository information, and actual biomarker instance data.

User manager implements different roles and appropriate permissions for each user role. It also manages data input/output and updates related to all user information in the system database.

User interface is the entry port for end users to interact with BiomRKRS. In the current version, the user interface implements access to all necessary functionalities of BiomRKRS, through appropriate user role and permission management, in a desktop application. The user interface has access to the other system components, namely, user manager, data manager, and query processor. The user interface component also has access to the BiomRKRS core ontology through the ontology management system, mainly to fetch vocabulary lists. At this stage, user queries are formed via direct selection from vocabularies inserted into the graphical components of the user interface instead of through natural language or free keyword-base search mechanisms. Figure 5 illustrates a snapshot of the user interface. A simpler interface has been planned to become available on the World Wide Web with more limited functionality available universally to individual users (subject to consideration of license agreements, especially for external knowledge resources incorporated into BiomRKRS).

2.4 System database

To store data pertaining to the different entities in BiomRKRS, a system database has been implemented. These data relate to:

- Ontologies: including all the data related to the physical location of the different external resources imported into the core ontology as well as their update history.
- System: including the physical location of the repository of the system for keeping local copies of related files.
- Biomarker instances: which include the actual biomarker data records. This also includes data on related pathways to each biomarker.
- Users: which includes all data to store for user roles, role permissions, and actual user instances registered in the system.

The system database is only directly accessible to the data manager and user manager components which then make it possible for the other parts of the system to have access to the system database.
Populating the BiomRKRS database

As mentioned earlier, BiomRKRS, as a biomarker information system, stores general (molecular) biomarker-related information but not patient-specific data. Therefore, there is a need for some mechanism to populate the database with information about identified (and potential) biomarker instances into the BiomRKRS biomarker database. Figure 6 shows the scenarios that have been identified as possible ways for gathering data and feeding them into this database.

The first possibility is to make use of other existing databases that report and store previously identified biomarkers, such as the databases shown on the left side of Figure 6, including OMIM (Hamosh, Scott, Amberger, Bocchini, & McKusick, 2005) and PharmGKB (Gong, Owen, Gor, Altman, & Klein, 2008). These individual data sources generally only cover a subset of contextual attributes that BiomRKRS defines as its core concepts and data features (see section 2.1 for the full list of attributes). Hence, BiomRKRS will serve as an integration platform, connecting information from diverse sources together.

The second option for populating the BiomRKRS database is through the use of textual documents in the biomarker-related literature. As we have demonstrated above, PubMed is a significant source of related research articles. In order to use these documents, one may use manual curation to extract and structure biomarker-related information. However, this does not scale to the massive amount of available literature. Therefore, we plan to make use of text mining techniques to automatically extract such information from the text of publications (or abstracts).

At this stage, our BiomRKRS biomarker database includes data that have been manually (by a domain expert) extracted and curated from the published literature related to two specific diseases, i.e., rheumatoid arthritis and osteoporosis.

BiomRKRS: technical specifications

The first version of the different functionalities of BiomRKRS has been implemented using Microsoft .Net 4.0 standard components. This includes the user interface component as well as all other core and behind-the-scene functionalities discussed in the previous sections.

Database access in BiomRKRS is through Language Integrated Query to Standard Query Language (LINQ to SQL), where the system database itself has been implemented in Microsoft SQL Server. Intermediate wrapper classes have been implemented around all database entities so that any change in the actual means and method of data storage will have a minimal impact on the other data-consumer components of BiomRKRS.

All ontology management and access in BiomRKRS has been implemented using Apache Jena Ontology APIs (Carroll et al., 2004) converted from Java to Microsoft C#.Net. Jena’s TDB triple store technology has been used for storing external large ontology files, such as LOINC’s instances. The RDFS reasoning schema implemented in Jena has been used to wrap the ontology model constructed using the ontology management component and to create an inference-ready extension of the ontology model. This extended ontology model is then used by the
knowledge reasoner component in BiomRKRS for inference purposes.

5 A clinical use-case

To demonstrate how BiomRKRS can assist clinical experts with their biomarker information search and related decision making process, we explore an example clinical scenario. In this case, Clinician has some prior and uncertain knowledge about some clinical measurements that have a role as an important prognostic biomarker for disease $d$ that is closely associated with a genetic pathway ($Pw$). Clinician is familiar with the terminology used in the ICD-10 disease classification and has worked with a variety of measurement methods as listed in LOINC’s method types. Clinician now decides to search the database of existing and known biomarkers for the following reasons:

- **Q1:** To understand whether there is any research and/or practical evidence to suggest an important biomarker for prognostic analysis of the genes involved in genetic pathway $Pw$ associated with $d$.
- **Q2:** If there are biomarkers found in answering Q1, to understand what the current validation status of a particular prognostic biomarker $PB$ is and as a result, whether she needs to carry out any further clinical assays in order to enhance her certainty about the effectiveness of $PB$; also what is the current status of $PB$, in general, for use by other clinicians or researchers.
- **Q3:** To understand whether further assays are necessary to confirm or enhance the validity of $PB$ as a prognostic biomarker for $d$, what methods of measurement have been used for $PB$ and on what sample (type). This will then lead to making the decision on which measurement method to use from the available methods to Clinician in her lab and how to plan her further study in regards to the subject patients she may have access to.

With these questions in mind, Clinician starts search with BiomRKRS. The process starts by finding the relevant disease term or category from the list of the ICD-10 diseases provided in BiomRKRS. After selecting disease $d$, BiomRKRS shows Clinician the list of associated genetic pathways to $d$. From the list of related pathways, she selects pathway $Pw$ and is now ready to further narrow down her search. BiomRKRS has fetched the list of clinical purposes from the external ontology QIBO (Buckler et al., 2013) and therefore, Clinician can select “prognosis” from the list of available purposes that BiomRKRS offers.

If Clinician knew the exact biomarker she was interested in finding further information about, she would avoid the selection and navigation process by simply searching for the exact title of biomarker $PB$ in BiomRKRS.

Supposing that Clinician does not specifically know any biomarker in this context, to find an answer to her first question (Q1), she now has access to the list of all prognostic biomarkers that relate to disease $d$ and pathway $Pw$ using BiomRKRS. Given that the BiomRKRS list of related biomarkers is not an empty list, Clinician is now able to look at all the evidence related to each data record (each representing a single biomarker) returned by BiomRKRS. Each data record corresponds to certain factors and conditions under which the specific biomarker has been measured, e.g., the patient population (from Experimental Factor Ontology), the measurement method type (from LOINC), molecular entity (from HGNC and HGVS), and validation status of the biomarker. Clinician finds links to the literature where there is evidence that specifically suggest $PB$ be a prognostic marker for $d$ as associated with $Pw$. She may also find some contradictory pieces of evidence that suggest the opposite be true for $PB$. She uses her own judgment, by looking at both sides and available study for each side, to decide whether to accept $PB$ as a biomarker for $d$ or not.

Having decided to accept positive evidence for $PB$ as a prognostic biomarker for $d$, Clinician decides to further her investigation and find an answer for Q2. She can see from the biomarker information record in BiomRKRS that $PB$ has not yet been fully approved by FDA as a qualified biomarker and its current status is “in research”. Although there are multiple research publications confirming $PB$ is a prognostic biomarker for $d$, there is no further evidence for it to have been fully investigated and validated against FDA rules and regulations. The validation status “in research” in BiomRKRS confirms that there is a need for more study/analysis in order for $PB$ to become truly “validated” and “qualified”. Clinician has now the answer for Q2 and is planning to take her investigation even further.

To answer Q3, Clinician looks at the data retrieved for $PB$, and in particular, the method type feature provided in
BiomRKRS information for PB. The fact that BiomRKRS retrieves multiple method types (techniques from LOINC) related to PB makes it possible for Clinician to assess the availability of tools in her lab and program a controlled assay for further investigating PB in the given context. Necessary samples and sample types (mapped to the SNOMED CT specimen concept) for conducting the trials is also a given piece of information from BiomRKRS. Clinician now has access to all information required to carry out further research and validation procedures on PB.

The knowledge reasoning feature of BiomRKRS plays an important role in this scenario in the situation where there is no study that has previously shown PB is a prognostic biomarker associated with genetic pathway Pw for disease d. BiomRKRS always returns not only the exact matches to Clinician’s constructed query, but also a list of biomarkers that semantically relate to her queries. In this case, BiomRKRS retrieves all prognostic biomarkers for d as well as for all diseases: \( \{d’ | d’ \in \text{non immediateParent}(d) \} \). This is mainly because by reasoning over the ICD-10 ontology, d is-a d'. Therefore, any biomarkers found for d’ may also apply to d. BiomRKRS ranks retrieved data records according to their relevance to Clinician’s queries. She will then decide whether to pursue an investigation specifically for PB in the context of disease d if there is evidence that shows PB has been identified as being a prognostic biomarker for a more abstract disease category, i.e., d’.

6 Planned future work

Development of BiomRKRS is an ongoing task; the current implementation includes all of the components introduced above in basic form. At this stage we have some concrete plans for further extensions and public deployment of the system to clinical experts and researchers. Among these action plans are:

- Implementation of full RDFS reasoning schema, using each inference mechanism for the core concepts from the BiomRKRS core ontology when semantically expanding end user queries.
- Designing and implementing an automated information extraction system for biomarker information extraction from free texts using text mining and natural language processing techniques. Currently all biomarker data in the system database have been extracted manually from related research publications. This text mining component will make it possible for BiomRKRS to store and have access to a large number of biomarker instances (on a larger set of diseases), the information of which can be associated to related textual documents, e.g., PubMed abstracts or full articles.
- Finding genetic mechanisms related to each disease (instead of or) in combination with pathway information. This will give the opportunity to end users to filter out biomarker data based on the participating genes in certain diseases.

7 Concluding remarks

With the current state of biological marker (biomarker) information resources and the drastic increase in the number of biomarkers identified for different clinical conditions and purposes, there is a need for effective and efficient information systems that can handle such a large and growing information-base. While there are already a number of systems available for biomarker data management and retrieval, there has been limited reuse of standard and existing vocabularies and none is capable of semantically processing user information requests. Our BiomRKRS system has been designed to overcome these two shortcomings with respect to other biomarker information systems, namely to make use of controlled vocabularies extracted via reuse of other well-established ontologies in the domain as well as to carry out reasoning over the constructed integrated ontology concepts of the system to semantically enhance user queries. As a result, BiomRKRS offers a biomarker data search procedure using controlled vocabulary terms and enables retrieval of exact matches as well as semantically related biomarker data. The semantic reasoning capabilities could potentially support the identification of new indications for previously unknown biomarkers related to certain clinical purposes.

8 Acknowledgements

In designing and developing BiomRKRS, we have received great support and knowledge from Professor John Wark at the University of Melbourne and Royal Melbourne Hospital. This work has been supported by National ICT Australia (NICTA) and the University of Melbourne. NICTA is funded by the Australian Government as represented by the Department of Broadband, Communications and the Digital Economy and the Australian Research Council through the ICT Centre of Excellence program.

9 References


Automatic Population of Structured Reports from Narrative Pathology Reports

YING OU, JON PATRICK
School of Information Technologies, The University of Sydney,
1 Cleveland Street, Sydney, NSW 2006
yiou6374@uni.sydney.edu.au, jonpat@it.usyd.edu.au

Abstract
The aim of this project is to use the methods of natural language processing to extract pertinent information from free-text pathology reports to automatically populate structured reports. A processing pipeline has been developed consisting of a combination of a supervised machine learning based approach using Conditional Random Fields for medical entity recognition and some rule-based methods. In total 477 narrative pathology reports of primary cutaneous melanomas were collected for evaluation.

Evaluations on the training set show that system performance can be improved by about 8.7% by refinement of the rules. The overall micro-averaged precision, recall and F-score of end to end evaluation on the test set are 89.44%, 80.60% and 84.79% respectively. Our study indicates the feasibility of this approach to automate the population of structured template from narrative reports with promising results. Error analysis reveals that a single specimen report with standard headings and the presence of simple and concise statements is significantly associated with correct populations.

In conclusion, the system can improve pathology reporting, and data mining for cancer registries, clinical audits and epidemiology research.

Keywords: Medical entity recognition, Structured output generation, Structured reports, Pathology.

1 Introduction
Generally, pathology reports can provide important information for the clinical management of cancer patients, allowing accurate diagnosis, staging and determination of prognosis. However, there can be several issues in traditional narrative reports compared to structured reports. For instance, essential elements are occasionally omitted, especially negative results are not always reported clearly. Recent research (Gomez and Tamboli 2005) has indicated that up to 27% of pathology reports in a cancer treatment centre failed to include at least one element of clinical significance, such as the status of vascular or lymphatic invasion. As well, the referring doctors often find it difficult to identify the necessary elements in a free-text pathology report to justify a given diagnosis.

There are a number of advantages for the use of structured pathology reports. For example, structured reports can improve the completeness of pathology reporting since they ensure that all information required are addressed and included in the report (Scolyer, Thompson, Stretch et al. 2004). The structured format has been proven to result in more complete reports for patients with breast cancer (Harvey, Sterrett, McEvoy et al. 2005). In addition, the information in a structured report is more predictable and easy to read which can improve the communication between pathologists and clinicians (Scolyer, Thompson, Stretch et al. 2005). One study (Thompson and Scolyer 2004) has indicated that the referring doctors find it easier to glean pertinent information from structured reports. Furthermore, structured reports can also facilitate efficient extraction of information for cancer registries, data collection and research purposes (Mohanthy, Piccoli, Devine et al. 2007).

As manual extraction of the information to populate a structured report is quite costly and time consuming, information extraction (IE) is preferable to automate the population process, as a sub-discipline of natural language processing (NLP) to extract significant information from unstructured data sources and transform it into structured data to facilitate access and retrieval of information. Named entity recognition (NER) is one of the sub-tasks of IE, aiming at identifying specific words or phrases (i.e., “entities”) and categorizing them. In the medical domain, the main themes of NER are medical entities, such as “clinical finding”, “body structure” and “procedure”. Many works have been reported for medical entity recognition (MER) tasks, wherein supervised machine learning based approaches have been widely adopted and achieved encouraging results (Patrick and Li 2010, Patrick, Nguyen, Wang et al. 2011, de Bruijn, Cherry, Kiritchenko et al. 2011, Torii, Wagholikar and Liu 2011).

Previous works on automatic structuring of free-text medical reports have attained some successes by classifying the relationships (e.g., dependencies) among medical entities with statistical methods at sentence level (Taira, Soderland and Jakobovits 2001). However, we believed that such approaches would not be a best fit for this project, since the structured report components required either very large segments of text that would be impossible to infer reliably, or inferences from the structure of a variety of medical entities to be properly
constructed. Hence, rule-based approaches are more suitable for automatic population of the structured output in this project.

The broad objective of this study is to automatically populate structured reports by using well-developed NLP-based methods. Specifically, we first aimed to identify medical entities with supervised machine learning based approaches, and then applied rule-based methods to construct the output for the report population with the correct entities.

2 Materials and Methods

2.1 Ethics Approval

The study protocol was approved by Royal Prince Alfred Hospital (RPAH), Sydney, Australia (Protocol No. X07-0322).

2.2 Materials

A collection of 477 prose pathology reports of primary cutaneous melanomas from patients referred to the Sydney Melanoma Unit at the RPAH in 2002 was drawn for this study. They were scanned, OCR-ed and de-identified.

We randomly selected 380 reports as a training set, 97 reports as a test set. Detailed annotation guidelines (see Appendix) have been developed according to Primary Cutaneous Melanoma Structured Reporting Protocol (1st Edition 2010) (Scolyer, Ellis, Heenan et al. 2010) by consulting pathologists. The training corpus was annotated manually with a text annotation tool (Visual Annotator (Wang 2009-2012)). The annotation process was carried out in a mixed conveyor method with a two phase validation and gold standard development. The annotation team was composed of 6 members divided into two groups. Each Group had a subset of the total tag set to annotate. Each team member annotated all files for those tags assigned to them. The team leader reviewed each annotation, as a validator for the development of the first gold standard. Although intra-annotator agreement cannot be evaluated in this method, high level of consistency can be attained.

2.3 Methods

The pipeline system is a combination of a traditional supervised machine learning based approach and some rule-based methods.

Figure 1 demonstrates the system architecture. From the diagram, raw records are passed through the pre-processing engine, which includes a sentence boundary detector and a tokenizer. A record is split into sentences, and then each sentence is split into tokens.

In a separate process, the training corpus is annotated manually to create gold-standards.

Subsequently errors in the manual annotations are identified by performing validation on the training data with a 100% train and test strategy (a recursive validation process: applies the computed model to validate the training data until no improvement of the performance can be made). With this self-validation process, more than 80 errors in concept annotation in the training data were detected. The errors were corrected manually so that the model would not learn from the incorrect examples. This process improved the scores by about 0.3%.

After pre-processing, the section context detector detects section context for each token as one of the features. In total five feature sets are prepared to train the Conditional Random Fields (CRF) (Sutton, McCallum and Rohanimanesh 2007) learners to identify the entities, and the outputs from the CRF are sent to the annotation file converter (Ann converter) to convert to the format that can be processed by Visual Annotator. Then the structured output generator populates the final outputs conformed to the structured templates.

2.3.1 Medical entity recognition

For MER experiments, the best model was obtained from five feature sets below:

1. Contextual features: a) Nine-word context window; b) Section context.

The medical category is the result of parsing the text to identify concepts of the Systematized Nomenclature of...
The ring-fencing tag is to identify chunks of a text with medical significance like scores or measurements by running trainable finite state automata over the text (Patrick and Sabbagh 2011).

3. Lexical features: a) Lemma, part of speech, chunk (obtained from the GENIA tagger (Tsuruoka 2004-2007)); b) Lowercase orthography of words; c) Expansions of abbreviations and acronyms; d) Correction of misspelling.


Affixes of length from 2 to 4 characters were used in separate experiments, which revealed that suffixes with a size of 2 and prefixes with a size of 3 produced the best performance. Moreover, nine-prefix context window yielded better performance than prefixes.

5. Bigram features: a) Bigram of words (or correction for misspelling).

2.3.2 Structured output generation

To populate the structured templates, a structured output generator was designed with rule-based methods, including processes described as follows:

1. Document classification

   At first, the documents needed to be divided into multiple specimen documents (documents contain more than one specimen) and single specimen documents (documents only contain one specimen).

2. Context Detection

   An engine was built to detect the section context information for each specimen.

3. Concept Extraction

   Concepts were extracted according to their types from the outputs of the Ann converter and then ranked with particular criteria: a potential candidate was assigned a salience measure based on a series of criteria (see Appendix), and the one with the highest salience measure was selected as the best candidate. Multiple post-processing modules were designed to extract values from the best candidate. It is worth mention that values for some fields require special rules to be extracted instead of ranking (e.g., “Cell growth”, “Previous melanoma”). The ranking criteria and post-processing modules were refined by evaluation against the gold-standard of training set and revised according to the error analysis.

4. XML Generation

   The XML generator generates the outputs in XML format from the extracted concepts based on the mapping strategies.

2.3.3 Statistical analysis

The performance of the system was measured by the standard Precision, Recall and F-score metrics, which are Precision = TP / (TP + FP), Recall = TP / (TP + FN), and F-score = (2* Precision* Recall) / (Precision + Recall), where TP—True Positive, FP—False Positive, and FN—False Negative.

3 Results

A baseline system was built using only the nine-word context window feature from the training corpus. The baseline system achieved the micro-averaged F-score of 78.95%. The full system was built using all features described above and achieved the best result of 84.29% F-score. Further experimental analysis of the contribution of feature types was conducted by progressively adding features to the system.

Table 1 presents the performance of the medical entity recognition experiments with 10 fold cross-validation on the training corpus.

<table>
<thead>
<tr>
<th>Model #</th>
<th>Features</th>
<th>Precision</th>
<th>Recall</th>
<th>F-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Nine-word context window</td>
<td>85.79%</td>
<td>73.12%</td>
<td>78.95%</td>
</tr>
<tr>
<td>2</td>
<td>M1 + Lowercase orthography of words</td>
<td>86.49%</td>
<td>77.12%</td>
<td>81.54%</td>
</tr>
<tr>
<td>3</td>
<td>M2 + Lemma</td>
<td>86.46%</td>
<td>78.21%</td>
<td>82.13%</td>
</tr>
<tr>
<td>4</td>
<td>M3 + Part of speech</td>
<td>86.12%</td>
<td>78.64%</td>
<td>82.21%</td>
</tr>
<tr>
<td>5</td>
<td>M4 + Chunk</td>
<td>86.07%</td>
<td>78.82%</td>
<td>82.29%</td>
</tr>
<tr>
<td>6</td>
<td>M5 + Medical category</td>
<td>86.35%</td>
<td>79.62%</td>
<td>82.85%</td>
</tr>
<tr>
<td>7</td>
<td>M6 + Expansions of abbreviations and acronyms</td>
<td>86.38%</td>
<td>79.64%</td>
<td>82.87%</td>
</tr>
<tr>
<td>8</td>
<td>M7 + Correction of misspelling</td>
<td>86.51%</td>
<td>80.16%</td>
<td>83.22%</td>
</tr>
<tr>
<td>9</td>
<td>M8 + Ring-fencing tag</td>
<td>86.41%</td>
<td>80.75%</td>
<td>83.48%</td>
</tr>
<tr>
<td>10</td>
<td>M9 + Suffixes</td>
<td>86.41%</td>
<td>81.15%</td>
<td>83.70%</td>
</tr>
<tr>
<td>11</td>
<td>M10 + Nine-prefix context window</td>
<td>86.55%</td>
<td>81.69%</td>
<td>84.05%</td>
</tr>
<tr>
<td>12</td>
<td>M11 + Section context</td>
<td>86.63%</td>
<td>81.75%</td>
<td>84.12%</td>
</tr>
<tr>
<td>13</td>
<td>M12 + Bigram of words</td>
<td>87.11%</td>
<td>81.65%</td>
<td>84.29%</td>
</tr>
</tbody>
</table>

Table 1: Scores for medical entity recognition experiments on the training set.

Firstly, we performed an evaluation on the structured outputs to establish the competence of the rules. Then another evaluation was carried out to reflect the improvement from refinement of the rules. Table 2 shows the results from structured output generation of first and second round of evaluations on the training corpus.

Finally, we used the best model to predict the test set, and then generated the structured outputs with the refined rules. Table 3 displays the results of end-to-end evaluation on the test set.

3.1 Error analysis

We performed manual inspection into each error and summarized the errors we could find explanations for into several categories (with estimated percentages).

3.1.1 First round of evaluation on the training corpus

In total 714 errors are identified in first round evaluation on the training corpus. The error types are:
Table 2: System performances (micro-averaged F-scores) from structured output generation of first and second round of evaluations on the training corpus.

1. Usability (26.05%)

A standard reporting convention and format is indispensable for data storage and retrieval. Inappropriate reporting format in some fields (particularly in “Clinical diagnosis” and “Subtype”) is the main reason for lower performance of those fields. For instance, “superficial spreading” is the standard reporting convention rather than “superficial spreading type” as the value for “Subtype” in “there is an intraepidermal element[“En:Primary Lesion”] of superficial spreading type[“Sy:Subtype”].”

2. Weakness in existing post-processing modules (24.09%)

In the initial design of post-processing modules, we failed to capture some useful information from the candidates. For example, “preexisting” is occasionally missed to be populated for “Assoc. benign naevus” in examples like “A preexisting benign dysplastic naevus [“En:Associated naevus (type) ”] is noted[“Li:Lexical Polarity Positive”].” The possible reason is that we use the results from GENIA tagger to determine the boundary of a noun phrase in a candidate, and in some cases, the tagger will tag “preexisting” as “VBG”, which is out of the scope of the noun phrase, hence it can not be populated correctly.

3. Incorrect annotations (20.73%)

Although during reflexive validation, most annotation errors have been identified and corrected, there are still a few errors detected in the first round of evaluation, accounting for the lowest F-score of “Cell growth” and “Prev. Rx / Trauma”.

4. Failure in mapping strategies (4.34%)

In the initial mapping strategies, we failed to map some annotation types to the associated fields, which is the major cause for the drop of F-score on “Other medical history”. For example, initially, we only mapped “En:Primary Lesion”, “En:Associated naevus (type)” and “De:Cosmetic Changes” to “Other medical history”, which led to occasional omission of some important information such as the history or duration of change (derived from “Li:Temporality”), size of lesion (derived from “De:Size”).

5. Other (21.85%)

Other reasons include encodings, weaknesses in exiting ranking criteria, etc. We assumed all characters in the reports are encoded in “utf-8”. However, due to mistakes of scanning or OCR, some characters are not encoded in “utf-8”, which lead to several invalid outputs.

The weaknesses in exiting ranking criteria comprise the conditions for applying the criteria and lexical entries of the criteria. For instance, some lexical entries like “partly” and “partial” (synonyms of “incomplete”) are omitted in “RegressCriterion”. Thus, “partly” are left out for “Regression” from “a partly regressed..."
3.1.2 Second round of evaluation on the training corpus

After the first round of evaluation, we took some measures to resolve the problems above, such as corrected annotations, replaced invalid characters with those encoded in “utf-8”, revised mapping strategies, modified existing ranking criteria and post-processing modules, designed new ranking criteria and post-processing modules, and represented output following standard convention and format.

The second round of evaluation revealed that most errors had been amended, and the total amount of errors reduced dramatically to 119. The modification or augment of ranking criteria and post-processing modules improved the performance significantly, but also had some adverse effect on the final outputs, accounting for 25.21% and 30.25% of the errors respectively. The second greatest contribution to the errors (20.17%) is from poor-writing of the original report, with some features like missing specimen identifiers and abnormal grammatical structures in some sentences. This issue is thought to be too difficult to resolve at present.

3.1.3 End to end evaluation on the test set

The micro-averaged F-scores in most fields are over 60%, except for “Desmoplasia”, “Distant metastasis”, “Excision margins: In-situ”, “Microsatellites”, “Other lesions” and “Previous melanoma”.

Generally, incorrect results from MER have prevented correction population in the first place, wherein the majority of which was due to the low recall of “En:Specimen Identifier”, which caused the drops of F-scores of most fields. The lowest F-scores are in the scarce fields (“Desmoplasia”, “Distant metastasis”, “Microsatellites”, “Other lesions” and “Previous melanoma”), due to a lack of sufficient training examples.

Ambiguity is another possible reason for the decrease of average F-scores on other fields, such as “Excision margins: Invasive” and “Excision margins: In-situ”. The instances of these related entity types “Ma:Excision In Situ” and “Ma:Excision Invasive” are so similar (e.g., “1.4mm from the closest lateral resection margin["Ma:Excision In Situ"]” and “1.6mm from the closest lateral resection margin["Ma:Excision Invasive"]”) that it is very difficult for the machine learner to discriminate them from each other, and the machine learner tends to misclassify the instances to the dominant type (which is “Ma:Excision Invasive” in this case).

The system has been released to the research community as a web page for testing data samples (please see http://www.icims.com.au/QUPPDemo for more details). Figure 2 illustrates an example of the final outputs for a single specimen document displayed in the web page.

4 DISCUSSION

4.1 Establishment of a specialized tag set

We established a specialized tag set based on the protocol and advice from pathologists. Preliminary study has shown that existing terminologies such as the Unified Medical Language System (UMLS) (Lindberg, Humphreys and McCray 1993) and SNOMED CT (IHTSDO 2007-2013) are not suitable for this task, as the concept categories they provide are too comprehensive and not tailored to the protocol, which cannot allow us to annotate enough of the text to complete a structured report.

4.2 Annotations on the pathology notes

It is noteworthy that pathology reports are distinguished from other clinical notes (e.g., compared to discharge summaries from the 2010 i2b2/VA Workshop on Natural Language Processing Challenges for Clinical

<table>
<thead>
<tr>
<th>Field</th>
<th>Number</th>
<th>Precision</th>
<th>Recall</th>
<th>F-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assoc. benign naevis</td>
<td>108</td>
<td>87.10%</td>
<td>84.38%</td>
<td>85.71%</td>
</tr>
<tr>
<td>Cell growth</td>
<td>108</td>
<td>74.14%</td>
<td>70.49%</td>
<td>72.27%</td>
</tr>
<tr>
<td>Clinical diagnosis</td>
<td>108</td>
<td>87.50%</td>
<td>80.77%</td>
<td>84.00%</td>
</tr>
<tr>
<td>Desmoplasia</td>
<td>108</td>
<td>66.67%</td>
<td>25.00%</td>
<td>36.36%</td>
</tr>
<tr>
<td>Diagnosis</td>
<td>108</td>
<td>93.88%</td>
<td>89.32%</td>
<td>91.54%</td>
</tr>
<tr>
<td>Distant metastasis</td>
<td>108</td>
<td>100.00%</td>
<td>16.67%</td>
<td>28.57%</td>
</tr>
<tr>
<td>Excision margins: Deep</td>
<td>108</td>
<td>100.00%</td>
<td>74.19%</td>
<td>85.19%</td>
</tr>
<tr>
<td>Excision margins: In-situ</td>
<td>108</td>
<td>80.00%</td>
<td>17.39%</td>
<td>28.57%</td>
</tr>
<tr>
<td>Excision margins: Invasive</td>
<td>108</td>
<td>69.84%</td>
<td>75.86%</td>
<td>72.73%</td>
</tr>
<tr>
<td>Level of invasion</td>
<td>108</td>
<td>95.18%</td>
<td>87.78%</td>
<td>91.33%</td>
</tr>
<tr>
<td>Lymphovascular invasion</td>
<td>108</td>
<td>95.83%</td>
<td>88.46%</td>
<td>92.00%</td>
</tr>
<tr>
<td>Microsatellites</td>
<td>108</td>
<td>66.67%</td>
<td>22.22%</td>
<td>33.33%</td>
</tr>
<tr>
<td>Mitotic rate</td>
<td>108</td>
<td>95.65%</td>
<td>83.02%</td>
<td>88.89%</td>
</tr>
<tr>
<td>Neuropotism</td>
<td>108</td>
<td>90.91%</td>
<td>81.08%</td>
<td>85.71%</td>
</tr>
<tr>
<td>Other lesions</td>
<td>108</td>
<td>100.00%</td>
<td>16.67%</td>
<td>28.57%</td>
</tr>
<tr>
<td>Other medical history</td>
<td>108</td>
<td>77.27%</td>
<td>66.67%</td>
<td>71.58%</td>
</tr>
<tr>
<td>Prev. Rx / Trauma</td>
<td>108</td>
<td>100.00%</td>
<td>44.44%</td>
<td>61.54%</td>
</tr>
<tr>
<td>Previous melanoma</td>
<td>108</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td>Regression</td>
<td>108</td>
<td>86.11%</td>
<td>83.78%</td>
<td>84.93%</td>
</tr>
<tr>
<td>Site and laterality</td>
<td>108</td>
<td>84.78%</td>
<td>83.87%</td>
<td>84.32%</td>
</tr>
<tr>
<td>Size of specimen</td>
<td>108</td>
<td>96.88%</td>
<td>90.29%</td>
<td>93.47%</td>
</tr>
<tr>
<td>Specimen type</td>
<td>108</td>
<td>89.90%</td>
<td>89.00%</td>
<td>89.45%</td>
</tr>
<tr>
<td>Subtype</td>
<td>108</td>
<td>92.65%</td>
<td>88.73%</td>
<td>90.65%</td>
</tr>
<tr>
<td>TILs</td>
<td>108</td>
<td>90.74%</td>
<td>85.96%</td>
<td>88.29%</td>
</tr>
<tr>
<td>TILs: Density</td>
<td>108</td>
<td>87.50%</td>
<td>58.33%</td>
<td>70.00%</td>
</tr>
<tr>
<td>TILs: Distribution</td>
<td>108</td>
<td>91.67%</td>
<td>61.11%</td>
<td>73.33%</td>
</tr>
<tr>
<td>Tumour thickness</td>
<td>108</td>
<td>94.32%</td>
<td>86.46%</td>
<td>90.22%</td>
</tr>
<tr>
<td>Ulceration(mm diam)</td>
<td>108</td>
<td>93.10%</td>
<td>85.71%</td>
<td>89.26%</td>
</tr>
<tr>
<td>Overall</td>
<td>3024</td>
<td>89.44%</td>
<td>80.60%</td>
<td>84.79%</td>
</tr>
</tbody>
</table>

Table 3: System performances (micro-averaged F-scores) for end-to-end evaluations on the test set.
Records (2010 i2b2 Challenges) (Uzuner, South, Shen et al. 2011), with following features:

1. Section headers. Section headers in a pathology report are more fixed, which can be summarized to six types: “Clinical History”, “Specimen Macroscopic”, “Microscopic”, “Diagnosis” and “Comment”; section headers in a discharge summary are more diverse, including “Chief Complaint”, “Past Medical History”, “Discharge Medications”, “Discharge Diagnosis”, etc.

2. Scope of a medical entity. Annotation guidelines of 2010 i2b2 Challenges pointed out that “Only complete noun phrases and adjective phrases should be marked”. Whereas, the scope of a concept in a pathology report can be more flexible, for example, a verb “shows” for “Li:Lexical Polarity Positive”, multiple noun phrases like “Breslow thickness 1.6mm” for “In:Breslow thickness(mm)”, a clause or sentence like “mitotic rate is 15 to 18 per mm2” for “De:Dermal mitoses”.

3. Focus of medical entity types. 2010 i2b2 Challenges mainly focused on three medical entities: “Problem”, “Test” and “Treatment”. Medical entity types of a pathology report can be more specific and detailed (e.g., “De:Site and laterality”, “De:Cell growth pattern”, “En:Primary Lesion” up to 41 types) in a melanoma pathology report. Thus, it requires more domain knowledge and training to annotate a pathology report.

4.3 Design of the structured template

The template has been slightly modified from that proposed in the protocol based on throughout analysis of the corpus. For example, the default unit for “Mitotic rate” is “per mm2” in the protocol. However, units like “per HDF (High Power Field)”, “per 5 HDFs” are also frequently used by pathologists in the corpus. Therefore, we finally decided to display the units aside from the numeric values to increase the flexibility to present this field in the template. Another example is the replacement of “Intraepidermal growth” with “Cell growth”, since in most documents, pathologists tended to describe growth patterns of different layers (e.g., dermis) of skin besides epidermis.

4.4 Comparison with other NLP systems

According to the protocol (Scolyer et al. 2010), a structured report should be “a report format which utilizes standard headings, definitions and nomenclature with required information.”

Our system performance is not compared to existing NLP systems like Medlee (Friedman, Shagina, Lussier et al. 2004) and cTakes (Garla, Lo Re, Dorey-Stein et al. 2011), since these systems have achieved relatively high performance on encoding of clinical documents or recognizing medical entities, at the cost of maintaining a
lexically variant-rich encoding table or dictionary, but our goal is different, that is to extract pertinent information from the free texts to populate structured templates rather than encoding.

### 4.5 Comments on the system

We demonstrated the feasibility of NLP-based methods that combined supervised machine learning based approach with some rule-based methods to automate the population of structured template. The pipeline system outcomes had achieved promising performance with overall averaged F-score of 84.79%.

Error analysis showed that a single specimen report with standard headings and the presence of simple and concise statements was significantly associated with correct populations. This is probably because:

1. The poorer performance of “En:Specimen Identifier” on MER could affect the final populations to a great extent. For example, “A.” can be presented as a block identifier rather than specimen identifier in some cases. If a specimen identifier is missed or misclassified, it could directly affect the results from document classification and finally negatively influence the outputs. We found that a good representation of a specimen identifier should either start with “Specimen” or include brackets (e.g., “Specimen A”, “(1)”).

2. As correct populations relied on correct detection of section headings in most cases, misuse or omission of section headings definitely hinder the final populations.

3. A simple and concise statement was also more likely to be detected by the NLP algorithms. For instance, it seems too difficult to populate correct values for “Excision margins: In-situ” and “Excision margins: Invasive” from “The nearest resection margins for the dysplastic junctional naevus[“En:A vessel “]”, “in situ” and invasive melanoma[“Sy:Diagnosis”] measure 1.8, 2.5 and 3.5nm respectively[“Ma:Excision Invasive”, “Ma:Excision In Situ”].” Firstly, current NLP tools like CRF cannot assign more than one tags for an overlap instance; secondly, even if the instances can be detected correctly, it still needs more complex rules to extract values from them.

The error analysis also addressed a critical and yet general issue that MER appeared to be the bottleneck of this project, which resulted in most of the errors. Some solutions may be useful to improve the system performance, such as exploring other features for better feature selection, ensemble multiple classifiers or machine learning algorithms.

### 4.6 Limitations

Our study was limited to a specific clinical domain and genre, so it might have a lack of generalizability or portability to some degree. Nevertheless, we believe that most of the features and rules adopted in the study are generalized, which can be applied to notes from other institutions. Moreover, some recommendations for writing melanoma pathology reports have been presented in a previous study (Patrick and Scolyer 2008), we hope pathologists can consider these recommendations if they want to use our system to process their notes.

We have displayed the structured outputs in a web page for pathologists to evaluate, but have been unsuccessful in recruiting any pathologists to engage fully with the task although a number volunteered initially. The extracts were validated by linguists trained to do this task. While they may not have been able to interpret the extractions as precisely as pathologists, their work has face validity and is internally consistent, which has been shown in our previous work on the project (Patrick and Scolyer 2008).

In view of the complexity and variability of language embedded in narrative reports, coupled with the existing error rate of our system, we believe that the system is worthy of further development.

### 5 Conclusion

We successfully developed an NLP-based system to extract pertinent information from narrative reports and automatically populate structured reports, and achieved promising results. We believe that the system can ensure significantly higher quality reporting for pathologists, as well as improve data mining for cancer registries, clinical audits and epidemiology research.

### Acknowledgments

The authors would like to give special thanks to other members of the Health Information Technologies Research Laboratory for their valuable contributions.

### Funding

This project is funded by the Quality Use of Pathology Program (QUPP) of the Department of Health and Ageing in Australia.

### 6 References


## 7 Appendix

### 7.1 Annotation guidelines

Below is the list of each annotation type with its brief definition (N.B. En: Entity, De: Descriptor, In: Invasion, Re: Reaction, Ma: Margins, Li: Linguistic, St: Structural, Sy: Synthesis).

- **En:** Associated naevus (type): Reference to any pre-existing or associated naevus with the melanoma. E.g., “a pre-existing naevus”, “previous naevus”.
- **En:** Primary Lesion: The primary lesion and typically the reason why the report is being prepared. E.g., “dermal component”, “lesion”, “nodule”.
- **En:** Lesion (other): Other lesions mentioned in the report and described. E.g., “second lesion”, “benign lesion”.
- **En:** Specimen Identifier: Reference to any satellites associated with the melanoma. E.g., “satellite”, “microsatellites”.
- **En:** Specimen Identifier: Used to identify the specimen. E.g., “specimen”, “sections”, “1.”, “1.”.
- **De:** Specimen type: The type of specimen that was used. E.g., “SNB”, “biopsy”, “ellipse of skin”.
- **De:** Cell Type: Description of the primary cell type. E.g., “epithelioid”, “spindled”, “naevoid, pleomorphic”.
- **De:** Cell growth pattern: The cell growth pattern contributes to the identification of the sub-type of melanoma. E.g., “pagetoid”, “lentiginous spread”, “vertical growth phase”.


“De: Cosmetic Changes”: Changes in appearance to the surrounding area that may be noted in the report and may be relevant to the diagnosis or the prognosis. E.g., “scar”, “necrosis”, “balloon cell change”

“De: Shape”: Description of the entity including colour, border and contour descriptions. E.g., “pigmented”, “purplish”, “brown”, “circular”, “linear”.

“De: Site and laterality”: The body part and side on which the lesion is located. This may also include finer locating information such as upper, lower, mid. E.g., “(R) neck”, “left thigh”, “lower back”, “mid back”.

“De: Size”: The sizes of the sample sent for testing, the primary lesion and any other lesions or notable entities. E.g., “2.5mm in diameter”, “10x3mm”, “25mm x 11mm x 6mm”.

“De: Ulceration (mm)”: Any reference to ulceration of a lesion or discontinuity of the skin resulting in loss of function. E.g., “ulcerated”, “ulceration”.

“De: Dermal mitoses”: Level of dermal mitosis and/or the mitotic rate. E.g., “occasional dermal mitoses”, “mitotic rate is 15 to 18 per mm²”, “mitotic activity numbers approximately 2 per 10 high power fields”.

“Ma: Excision Clear”: Statement that the excision margins are clear or any descriptive material relating to the excision margins that doesn’t belong under the other tags. E.g., “excision appears complete”, “lines of excision are clear”.

“Ma: Excision deep”: The distance from the lesion to the deep margin. E.g., “deep margin of 2.3mm”, “Distance from deep margin = 4.0mm”.

“Ma: Excision in situ”: The distance from the in-situ or junctional component to the lateral margins. E.g., “closest peripheral margin to in situ component measures 0.9mm”, “in situ component is situated 0.2mm”.

“Ma: Excision invasive”: The distance from the dermal component to the lateral margins (default category for lateral margins). E.g., “closest peripheral margin to invasive component = 1.0mm”, “5.95mm to the nearest lateral margin”.

“Re: Desmoplasia”: Reference to the presence or absence of desmoplasia. E.g., “desmoplastic reaction”, “desmoplasia”.

“Re: TILS”: Any reference to tumour infiltrating lymphocytes (TILs). E.g., “lymphocytic infiltrate”, “tumour infiltrating lymphocytes”.

“Re: Solar Elastosis”: Evidence of skin reaction to sun. It will make the skin appear leathery and can impact on diagnosis and prognosis. E.g., “solar elastosis”, “sun-damaged skin”.

“Re: Fibrosis”: Evidence of reaction in connective tissue or scarring. E.g., “fibrosis”, “angiofibroplasia”.

“In: Vascular/lymphatic invasion”: Any reference to infiltration of the blood vessels and lymphatic system. E.g., “vascular invasion”, “vascular/lymphatic invasion”.

“In: Neurotropism”: Reference to any neurotropism present or absent. E.g., “perineural invasion”, “neurotropism”.

“In: Breslow thickness (mm)”: Primary tumour thickness. E.g., “Breslow thickness 1.6mm”, “depth 0.55mm”.

“In: Clark level”: The layer of the skin into which the tumour has permeated. E.g., “Clark level 4”, “Clark level II”.

“Li: Lexical Polarity Negative”: Lexically bound (either in a separate lexical item or as a suffix or prefix) polarity that is modifying something else. It is related to negation. E.g., “no”, “not”, “lack”, “nor”.

“Li: Lexical Polarity Positive”: Lexically bound (either in a separate lexical item or as a suffix or prefix) polarity that is modifying something else. It is related to confirmation. E.g., “shows”, “evidence”, “noted”, “present”.

“Li: Modality”: Modality in the positive direction. E.g., “possibly”, “probably”, “may”, “definite”.

“Li: Mood and Comment Adjuncts”: Indication of degree or intensity. E.g., “mild”, “moderate”, “some”, “small”.

“Li: Temporality”: Any reference to temporal indicators. E.g., “previous”, “early”, “late”.

“Sy: Diagnosis”: The diagnosis of the lesion within the specimen. E.g., “malignant melanoma”, “melanoma”, “basal cell carcinoma”.

“Sy: Regression”: Any reference to regression within the lesion. E.g., “regression”, “regressive activity”.

“Sy:_subtype”: Refers to the histological type and classification of the melanoma only and will typically map onto the diagnosis or in descriptions of the lesion. E.g., “superficial spreading type”, “nodular”, “lentigo maligna”.

“St: Clinical History Heading”: Any heading that pertains to the history of the patient (or the specimen if no specimen heading included). E.g., “CLINICAL NOTES”, “CLINICAL DETAILS”, “CLINICAL HISTORY”.

“St: Specimen Heading”: Any heading that pertains to the Specimen. E.g., “SPECIMEN”, “NATURE OF SPECIMEN”, “Specimen(s) Received”.

“St: Macroscopic Heading”: Any heading that pertains to the Macroscopic Examination. E.g., “MACROSCOPIC”, “MACROSCOPIC...”
EXAMINATION”, “MACROSCOPIC REPORT”.
“St:Microscopic Heading”: Any heading that pertains to the Microscopic Examination. E.g., “MICROSCOPIC”, “MICROSCOPIC REPORT”, “MICROSCOPIC EXAMINATION”.
“St:Diagnosis Heading”: Any heading that pertains to the diagnostic summary. E.g., “DIAGNOSIS”, “SUMMARY”, “CONCLUSIONS”.
“St:Comment Heading”: Any heading that pertains to comments made by the pathologists. E.g., “Comment”, “COMMENT”, “Further report”.
“St:Sub Heading”: Any miscellaneous subheading that does not fall under the aforementioned structural headings tags. E.g., “Growth pattern”, “Lines of Excision”, “Cytological features”.

7.2 Criteria for ranking candidate concepts
The ranking criteria for concept extraction are described as follows:

1. **LengthCriterion**: Returns 1, if the candidate is the longest candidate; else 0.
2. **UppercaseCriterion**: Returns 1, if the candidate is uppercase; else 0.
3. **PolarityCriterion**: Returns 1, if instance(s) of “Li:Lexical Polarity Negative” or “Li:Lexical polarity Positive” and the candidate occur in the same sentence for most fields; -1, if instance(s) of “Li:Lexical Polarity Negative” and the candidate occur in the same sentence for “Diagnosis” and “Subtype”; else 0.
4. **MeasurementCriterion**: Returns 1, if there is a token inside the candidate ends with “mm” as its unit; else 0.
5. **ClearCriterion**: Returns 1, if instance(s) of “Ma:Excision Clear” and the candidate occur in the same sentence; else 0.
6. **FrequencyCriterion**: Returns variable values (the frequencies of the tokens inside the candidate in other candidates), if the tokens inside it occur in other candidates; else 0.
7. **PrimaryCriterion**: Returns 1, if the candidate refers to the primary lesion; else 0.
8. **ModalityCriterion**: Returns 1, if instance(s) of “Li:Modality” and the candidate occur in the same sentence and its polarity undetermined; else 0.
9. **TemporalityCriterion**: Returns 1, if the candidate is a “Sy:Regression” instance, “Li: Temporality” instance(s) and the candidate occur in the same sentence; -1, if the candidate is a “De:Specimen Type” instance, “Li: Temporality” instance(s) and the candidate occur in the same sentence else 0.
10. **PositiveCriterion**: Returns 1, if the polarity of the candidate is “positive”; else 0.
11. **MarginCriterion**: Returns 1, if the candidate is a “Ma:Excision Clear” instance with any instance (s) of “Ma:Excision In Situ”, “Ma:Excision Invasive”, or “Ma:Excision Deep” in the same sentence; else 0.
12. **DistributionDensityCriterion**: Returns 2, if the candidate or “Li:Mood and Comment Adjuncts” instances around it contains lexicon(s) indicating both distribution and density; 1, if only refer to distribution or density; else 0.
13. **AcronymCriterion**: Returns -1, if the candidate is an acronym; else 0.
14. **MarginTypeCriterion**: Returns 1, if the candidate is a “Ma:Excision Clear” instance which indicate the type of excision margins (in-situ, invasive, deep); else 0.
15. **BodyStructureCriterion**: Returns 1, if the medical category of the candidate is “Body structure”; else 0.
16. **LateralityCriterion**: Returns 1, if the candidate contains laterality (e.g., “left”, “right”); else 0.
17. **MelanomaCriterion**: Returns 2, if the candidate is a “melanoma”; 1, if it indicates any malignancy; else 0.
18. **NaevusTypeCriterion**: Returns 1, if the candidate indicates the type of naevus; else 0.
19. **LevelCriterion**: Returns 1, if the candidate states the classification of Clark level; else 0.
20. **RegressCriterion**: Returns 1, if the candidate indicates stage or characteristic of regression; else 0.
21. **DimensionCriterion**: Returns 1, if the candidate is a size of two or three dimensions; else 0.
22. **PositionCriterion**: Returns 1, if the candidate is in the first position; else 0.
23. **SpecimenTypeCriterion**: Returns 1, if instance(s) of “De:Specimen type” occur nearby the candidate; else 0.
24. **RateCriterion**: Returns 1, if there is a token inside the candidate ends with units like “per mm 2” or “per hdf”; else 0.
25. **InvasiveCriterion**: Returns 1, if the candidate indicates an invasive lesion; else 0.
26. **MoodDegreeCriterion**: Returns variable values (from 0.5 to 3), depends on degree or intensity of “Li:Mood and Comment Adjuncts” instance(s) around the candidate; else 0.
27. **DiagnosisCriterion**: Returns 2, if a “Sy:Diagnosis” instance precedes or follows the candidate; 1, if they are in the same sentence; else 0.
28. **BreslowCriterion**: Returns 1, if the candidate indicates it is a “Breslow thickness”; else 0.
What Pacific people think of online mental health information

BERNADETTE PENI, KAREN DAY, MARTIN ORR
School of Population Health, Faculty of Medical and Health Sciences
The University of Auckland
Private Bag 92019, Auckland 1142, New Zealand
k.day@auckland.ac.nz

Abstract
The purpose of this research was to examine how the attitudes and beliefs of Pacific people impact on their propensity to use online mental health service www.depression.org. The mental health of the Pacific population in New Zealand is concerning - there is a real need to address this health concern. One method of doing so is through online mental health information for support and self-help. Little has been documented about the attitudes, perceptions and propensity of Pacific people to use online mental health information.

Data was collected through means of a questionnaire and a paired discussion with Pacific participants. The findings revealed mixed opinions amongst the participants about their view of online mental health information due to place of birth, experiences and language. We conclude that the attitudes, beliefs and propensity of Pacific people to use online health resources are dependent on their upbringing, experiences and cultural values. This research is exploratory and therefore more research is required to gain a reflective understanding of how Pacific people view online mental health resources.

Keywords: mental health, Pacific, culture, etherapy.

1 Introduction
According to the first mental health survey in New Zealand in 2006, Pacific people have been reported to experience mental health disorders at higher levels compared to the general population (Browne et al., 2006, Agnew et al., 2004). With the researcher being of Samoan ethnicity, it was with great interest and curiosity that the topic of e-mental health be looked into to gain a Pacific perspective. Pacific people appear less likely to use mental health services and make mental health visits in comparison with the general population (Browne et al., 2006). Although there may be varying issues behind the underutilisation of mental health services, little has been documented about Pacific people and the usage of online mental health information and/or services (Finau, 1994).

A major benefit of online mental health information is that it allows one to overcome geographical barriers and access health information/services more readily (Cline and Haynes, 2001). The Internet allows the search for mental health information in the privacy and discretion of an individuals home or other private spaces. (Cline and Haynes, 2001, Griffiths et al., 2006, Wellman and Haythornthwaite, 2002). If people are able to easily access mental health information and/or services online and privately, then are Pacific people doing so and what obstacles do they face? The purpose of this research was to examine how the attitudes and beliefs of Pacific people impact their propensity to use an online mental health service.

The website chosen for our research was www.depression.org.nz (see Figure 1) fronted by Sir John Kirwan, a well-known All Black, and has himself experienced depression (Kirwan and Thomson, 2010). The site is an interactive online service based on cognitive behavioural therapy, which is widely used for the treatment of depression and other mental illnesses. Currently, four types of mental health interventions exist that can be delivered online (Barak and Grohol, 2011):
1. Psycho-educational sites providing information only
2. Online counselling and therapy, where the patient and therapist use the Internet to communicate
3. Online self-help support groups and blogs

An interactive self-guided intervention is defined as ‘a technology, most often a website, that offers an individual the opportunity to interact with a structured, self-guided software program online that steps them through a program of self-help. These programs are usually drawn from the cognitive-behavioral literature and offer interactive exercises to the user.’ (Barak and Grohol, 2011)[p.18]. We chose The Journal (website) as an example for Pacific people to review in order to examine how their attitudes and beliefs impact their propensity to use online mental health services.

Figure 1: Screengrab of www.depression.org.nz
The current form of the website was launched in June 2010 as a central component of the New Zealand National Depression Initiative. This is a multimedia campaign that creates awareness and understanding of depression and guides people towards the website for help. The website provides further in-depth information and education about depression and available services. The website also has an integrated self-management program that users can sign up for called ‘the Journal’. The Journal guides users through a sequence of evidence based lifestyle changes, and aims to develop problem solving skills. The Journal includes a range of expert advice interviews and integrated animation. In addition John Kirwan acts as a coach or guide connecting each component and facilitating the discussion of key issues with psychiatrist and psychologist experts. An integrated clinical support component Ia for those who wish to use it by clicking on the link provided. The Journal contains many of the components of e-therapy but is described as an e-learning programme; it aims to complement other services and forms part of a broader integrated stepped care model.

2 Methods

Since this was an exploratory study, we chose to use a qualitative approach to seek the ‘why’ and ‘how’ aspects of the topic (Denzin and Lincoln, 2005), and therefore build a richer picture surrounding the research question. This approach provides opportunities to access in-depth conversations and to gain insights and information from participants (Silverman, 2001). Ethics approval was granted on 2 November 2011 by the Northern Y Regional Ethics Committee (Reference Number: NTY/11/EXP/061). The research was conducted as part of the first author’s Masters thesis. Participants were recruited and data gathered over a three month period in 2011.

The overall research project’s methods consisted of a systematic literature review on mental health, e-therapy, online health information, and Pacific people; a set of interviews with five key informants; and interaction with four Pacific island volunteer participants who completed a questionnaire and interacted with the researcher in paired interviews (described below). The key informants included a consultant psychiatrist (specialist in Pacific islander mental health), two general practitioners from primary care, the Chief Executive Officer and Director of Interaction (social marketing and design) of two of the organisations that established the website. For the purpose of this article, we are reporting on the data collected from the four participants’ questionnaires and interviews.

To be included in the study, participants were required to meet all of the following eligibility criteria:

- Pacific ethnicity
- 24 – 45 years of age
- Able to read, speak and understand English to participate in the survey and understand the content presented for paired discussion with the researcher (first author)
- Not already be diagnosed with a mental illness as an ethical concern
- Be available for approximately 1.5 hours to complete the questionnaire and participate in the paired interviews.

Initial attempts at ‘conventional’ recruitment activities were not productive, e.g. an advertisement on notice boards at churches and word of mouth recruitment through the researcher’s network. A more directive approach was used in which an invitation letter was distributed to the friends and family members of personal friends through an acquaintance of the researcher (first author). The purpose of sending the letter through an acquaintance was to overcome the ‘friend get friend’ power and the likelihood that the researcher knew a participant personally. People who were willing to participate contacted the researcher and were screened for eligibility. Once it was established that a participant met the eligibility criteria and was still willing to participate, a convenient time for the participant to meet up at their house or location of convenience was established to complete the questionnaire and the interview. Due to research time constraints, of the ten invites, four participants volunteered to take part. It is most likely however that more participants would have taken part if more time was permitted.

The data collection consisted of the following steps at the appointment time:

1. A brief introduction that included the purpose of the study, what participating in the study would involve, and an opportunity for the participants to ask questions.
2. Upon agreement to take part in the research, participants signed the consent form. The consent form was signed to ensure that each participant agreed with terms such as the storage of their information for ten years, the interview being digitally voice recorded and the delivery of information in this study to be adjusted for anonymity.
3. The questionnaire was completed by each individual.
4. The researcher played a television advert from YouTube (DraftFCB NZ, 2010) about the website, featuring Sir John Kirwan.
5. Participants were shown the website, www.depression.org.nz and invited to browse through it in preparation for the interviews.
6. Participants were interviewed one by one with the researcher.
7. Once the interviews/discussions were completed, the participants were thanked and the appointment drawn to a close.

2.1 Questionnaire

The questionnaire consisted of 21 questions. The structure of the questionnaire comprised four sections:

- General Background, to capture the details of the participant such as age, gender and ethnicity.
- Propensity to use online information services, to build a picture of the participants’ knowledge about the Internet and computers or laptops, e.g. frequency of use and terminology used or understood. This was to provide a picture of what it may be like for most Pacific people in
New Zealand and the possible reasons why Internet and computers are used or not used.

- **Attitudes**, to provide a picture of the participants’ attitudes and behaviour towards seeking health information for either themselves or others.
- **Beliefs**, to provide insight into the beliefs of the participants’ ethnic group and how mental illness is interpreted.

Responses were recorded on a five point Likert scale, e.g. ranging from ‘very comfortable’ to ‘not comfortable at all’. The data were analysed by counting them up and creating tables to compare and contrast answers to groups of questions.

### 2.2 Interview

A semi-structured interview was conducted once the participants had completed the questionnaire, watched the advert and browsed through the website. Before participants were left to navigate and look through the website, the researcher explained what the ‘my journal’ feature was (as displayed in bottom left hand corner of Figure 1). Participants were asked not to look into this feature as this required them to fill out the patient health questionnaire (PHQ-9) ‘self-test’, subsequently creating a personalised journal for depression management. The website provides therapeutic routines and activities, e.g. cooking and exercise, as part of a broader cognitive behavioural therapy programme specifically designed for online self-directed delivery.

To avoid any suggestion that the participants themselves may be suffering from mental health, the researcher described what ‘my journal’ contained and asked them not to explore it during the interview session. Participants were otherwise welcome to investigate further if they wanted to, after the interview ended. Following this discussion, each participant was left for a few minutes to navigate and read through the other features of the website. Participants were asked to alert the researcher when they were satisfied they had browsed through the website for long enough. Upon this signal, the researcher conducted the semi-structured interviews that were digitally recorded. Interviews lasted approximately 45 minutes to an hour.

The interviews were analysed thematically and compared and contrasted with the questionnaire analysis.

### 3 Findings

The findings of the questionnaire and interviews are presented separately.

#### 3.1 The questionnaire

All participants were female and were in the age brackets between 24-30 or 41–45 years of age. Of the four women, two were of Samoan ethnicity, one Niuean and one Tongan ethnicity. Three of the four participants were currently employed whilst the fourth identified themselves as a ‘full time student’. Two women specified that they were born in the island of their ethnic groups, indicating migration to New Zealand while the remaining two were ‘New Zealand born’.

No-one was suffering from a chronic illness. All participants indicated that they had access to the Internet and owned either a computer or laptop. Notably, each participant indicated they had completed either secondary school or tertiary education. Although all four participants owned a mobile phone, only two had the feature of Internet access available on their phone. Neither of them had ever used the Internet feature on their phones to access health information.

#### 3.1.1 Propensity to use online services

Only three out of the four women noted that they were ‘very comfortable’ with using a laptop or computer on their own and required no assistance from someone else. These individuals knew how to operate the device, create documents, to write a letter and to connect to the Internet. In contrast, the fourth participant indicated that she was not comfortable using a computer despite secondary schooling in New Zealand and owning a computer or laptop. When asked the question, ‘How comfortable are you using the Internet?’ only two answered that they were very comfortable using the Internet and were familiar with aspects of the Internet such as using a search engine and emailing. One specified that she was not confident in her skills of using the Internet whilst the fourth indicated that she was not completely confident in her usage and knowledge of the Internet. The usage of the Internet per week by participants varied from a rating of 3 (5-6 times) to 5 (8-10 times) whereas the usage of a computer or laptop outside the hours and facility of work places varied from a rating of 2 (3-4 times) to 5 (8-10 times).

When asked ‘Have you ever used the Internet to search for health information for yourself?’ two participants answered ‘no’ whilst the other two answered ‘yes’. Of the two who answered ‘yes’ one said, ‘I think it was two weeks ago,’ and the other, ‘A year ago’. The nature of the search was not specified by either participant. Likewise, when answering the question ‘Have you ever used the Internet to search for health information for another person?’ two answered ‘no’. The remaining two responded ‘yes’ with the following descriptions of ‘Checking a suitable rest home within South Auckland for an elderly relative,’ and, ‘To research a relative’s ailments/symptoms and the medication diagnosed to them from their GP’. These answers indicated a difference in how one interpreted or understood ‘health information’ to be. The question ‘What do you think a ‘trusted website’ is and why?’ was met with the following answers:

- ‘One that is based on evidence i.e. relevant research: research in numbers. It is a secured website i.e. has the lock symbol in the web address.’ (A)
- ‘Security with my information which is very important or I am hesitant to use it.’ (B)
- ‘Trusted website is one with quality health information given by legitimate health professionals.’ (C)
- ‘A website that has credibility towards it of some sort, that the public can trust to use, that it is not made created on falsified facts.’ (D)

These answers show how differently people interpret the word ‘trust’ with regard to Internet websites. When asked, ‘Have you ever used the Internet to search for online mental health information?’ only one participant responded with ‘yes’ stating they were ‘curious about the ad with ex-All Blacks player John Kirwan on depression’. Interestingly, this answer was provided
before the participant was asked to look at the website or any mention of the depression campaign fronted by Sir John Kirwan. Participants were then asked to answer three questions about attitudes.

3.1.2 Attitudes

This section covered questions that were related to the behaviour and attitudes participants had towards seeking health information for either themselves or others (family members or friends). Tables 1 and 2 show the responses to these questions.

<table>
<thead>
<tr>
<th>Participant</th>
<th>Rating</th>
<th>Reason for rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>5</td>
<td>Because I know the importance of information and know there is no harm in taking a look.</td>
</tr>
<tr>
<td>B</td>
<td>5</td>
<td>I would like to know more information of what services are out there to cater for my condition. Also to know more about the condition that I have been diagnosed with.</td>
</tr>
<tr>
<td>C</td>
<td>3</td>
<td>Having it from a doctor themselves suits me best.</td>
</tr>
<tr>
<td>D</td>
<td>3</td>
<td>I chose the ‘maybe’ option because if someone tells me to check stuff out online about health, I probably wouldn’t really care and won’t check it out. But then I might because I’m so used to being on the Internet all the time. So, yeah, half and half.</td>
</tr>
</tbody>
</table>

Table 1: Summary of Participant Responses about attitude to using a health website

<table>
<thead>
<tr>
<th>Participant</th>
<th>Sources participants turned to for health information</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>GP clinic (GP or nurse)</td>
</tr>
<tr>
<td></td>
<td>Relative – she’s a nurse (RN)</td>
</tr>
<tr>
<td>B</td>
<td>Doctor</td>
</tr>
<tr>
<td></td>
<td>Word of mouth (friends, relative, church…)</td>
</tr>
<tr>
<td></td>
<td>Internet</td>
</tr>
<tr>
<td></td>
<td>TV</td>
</tr>
<tr>
<td>C</td>
<td>Hospitals</td>
</tr>
<tr>
<td></td>
<td>Clinics</td>
</tr>
<tr>
<td>D</td>
<td>The doctors</td>
</tr>
<tr>
<td></td>
<td>My parents (my dad used to practise a lot of medical-related treatment – so I get most of my knowledge from him.)</td>
</tr>
</tbody>
</table>

Table 2: Summary of participant responses about sources of health information

The first question asked, ‘When encouraged by others for example, doctors, specialists, friends and family, to use a health website, how likely are you to follow through?’ Participants were asked to circle their answer on a Likert scale and then indicate why they selected that answer. The scale ranged from 1 (very unlikely) to 5 (most likely). Although two participants circled 3 for ‘maybe’ and the other two circled 5 for ‘most likely’ each gave a different answer (see Table 1). The answers varied from an attitude of ‘don’t care’ to a preference to see a doctor. Of the two who answered with a 5, these participants had indicated that they did search for health information either for themselves or for a friend in the previous section of the questionnaire, in the computer and knowledge section.

The next question was asked to capture feedback concerning where they sourced health information. The most common responses amongst all four were either ‘doctors’, ‘clinic’ or ‘hospital’ (Table 2).

In the list provided by the participants only one mentioned the Internet. This aligned with their previous answers - they did look up health information for another person and rated they were ‘most likely’ to use a health website if informed by another. Though there was no mention of these in the questionnaire, two responses included family members as sources of health information as they were currently working within the health sector or used to work in a health-related position.

The final set of questions concerned the influence of beliefs on the use of online health information sources.

3.1.3 Beliefs

The questions asked in this section related to the individuals’ ethnic group and how mental illness is interpreted within this group. Only two questions were asked. The first question was met with similar responses. This similarity relates to the issue of stigma or discrimination still being present within the ethnic groups the with which participants identified themselves (Samoan, Tongan and Niuean). Table 3 presents the answers, showing how the issues of stigma and discrimination are still present within ethnic groups with which the participants identified.

<table>
<thead>
<tr>
<th>Participant</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Traditionally quite a taboo or religious phenomenon. But I think it is becoming a more common topic that is being talked about among Samoan people, mainly due to more awareness about mental illness – adopting Western views on it.</td>
</tr>
<tr>
<td>B</td>
<td>Lack of knowledge. In my upbringing we were told that people with mental illnesses’ were ‘crazy/naughty’ and ‘difficult’. Definitely not to associate ourselves with them whether they are a relative or very close family friend.</td>
</tr>
<tr>
<td>C</td>
<td>Mental health is not really talked about. Those with a mental illness are usually left to do their own thing. Not really shunned by the family but also not embraced.</td>
</tr>
<tr>
<td>D</td>
<td>In my Tongan culture, I think and believe that it is a taboo to even talk about mental illness. So therefore I don’t hear it much – as in there’s no talk that I know of which has to do with mental health. The only times that I do hear about it is when someone is talking about ghostly stories back in the Islands. This is called ‘puke faka tevolo’ which means the devil’s sickness. Its typically seen as crazy and not part of society.</td>
</tr>
</tbody>
</table>

Table 3: Responses that explain how mental illness is explained, viewed or talked about

One participant acknowledged that these views are slowly changing due to New Zealand’s effort to ‘normalise’
mental health and raise awareness. Participants were asked, ‘Based on the views and beliefs of your ethnic group, how is mental illness explained, viewed or talked about?’ The responses in Table 3 indicate that mental illness has negative connotations for Pacific ethnic groups. In other words, the issue of mental health problems is often reflected and portrayed in a negative light. It is of no surprise then, that Pacific Islanders hesitate or refuse to seek help until they are too ill (e.g. severe depression). When such an outlook against mental health issues persists, it is no wonder that stigma and discrimination continue to be obstacles to overcome within the Pacific population.

The next question asked was ‘What words are used by your ethnic group to explain or understand depression and or mental illness?’

This question was asked to provide an indication of what words were used in the participants’ ethnic languages to communicate the issue of mental health or depression. Table 4 summarises the responses to this question. The most common term or understanding was ‘crazy’.

<table>
<thead>
<tr>
<th>Ethnic group</th>
<th>Word</th>
<th>Participants’ understanding or meaning of the word</th>
</tr>
</thead>
<tbody>
<tr>
<td>Samoan</td>
<td>Ma’i</td>
<td>= Sick</td>
</tr>
<tr>
<td></td>
<td>Leaga le ulu</td>
<td>= Crazy or sick in the head</td>
</tr>
<tr>
<td></td>
<td>Valea</td>
<td>= Crazy, dumb</td>
</tr>
<tr>
<td>Tongan</td>
<td>Puke faka</td>
<td>= Devil’s sickness (refers to the mind sickness)</td>
</tr>
<tr>
<td></td>
<td>tevolo</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Fakasesele</td>
<td>= Crazy</td>
</tr>
<tr>
<td></td>
<td>Siasi</td>
<td>= Crazy</td>
</tr>
<tr>
<td>English</td>
<td>Crazy</td>
<td>= Dumb in the head – crazy</td>
</tr>
</tbody>
</table>

Table 4: Summary of responses about ethnic language for mental illness

The participant of Niuean descent noted that she could not think of any words but, because she was married to a Samoan man, she knew one Samoan word commonly used to describe the mentally ill. Another participant noted that they could not remember specific words or the spelling of Samoan words. Instead she stated she knew the English equivalents and provided these. As a result only Samoan and Tongan words were gathered (with the addition of the English equivalent as provided by one participant). Notably the word ‘crazy’ appears to be a common word to explain depression or mental illness. When the participants indicated that they had completed the questionnaire the session moved to the paired discussion.

3.2 Paired Discussion

Once the participants had viewed the advert and navigated through the website (minus the ‘my journal’ feature), a semi-structured interview took place between the researcher and the participant. This paired discussion was based on questions from the questionnaire and in relation to the website. For example, questions included ‘As a Pacific Island individual, what are your beliefs about using online health information?’ and ‘What features of online health information would you like to change or improve to increase the likelihood of your using online health information?’ The purpose of repeating questions from the questionnaire was to allow room for further explanation, because participants may not have had the patience to write a meaningful response, may have provided broad statements or points, or did not answer in writing. The paired discussion meant that participants had the opportunity to tell stories and elaborate on what they may have already provided. This allowed for a richer picture to develop and provided the opportunity to capture information that may not have been included in the questionnaire.

Before the researcher began the paired discussion, some of the participants made comments after they were satisfied with looking at the website and its features. Although all participants were familiar with the ‘John Kirwan’ adverts about depression and website, some found the website as ‘something new’ they had not looked into or considered, as expressed by one participant.

That was interesting… I had seen it on TV but this is actually the first time I’ve looked into it. (A)

Some participants felt that the interface for the website was ‘good’ and ‘easy to follow’. This was because the website interface was clean (clearly labelled), not cluttered with tabs and had an ‘inviting’ feel with an image of a park bench. Being able to navigate easily through a well-labelled and simple website proved to be a helpful feature that encouraged them to use online mental health information. This is illustrated through the following responses.

‘Navigation in the website is good. The tabs and headings are pretty good, and it kinda flows from left to right. The tabs take you from the info to the resources and content...an easy logical flow – you don’t get confused on where to go and its clearly labelled.’ (D)

‘It’s really easy to follow and very simple. Initially it looks really good, you just go off the first page just scrolling through it. I think for me it will depend on my mood because there’s quite a lot of reading to go through, but, it would definitely be good for me because it explains things in more detail and has the options to watch a video; it’s not just reading, it’s about those other options. I think more so the interactive part I would be more interested in.’ (B)

‘It’s easy, it starts off describing what depression is, then the causes, it goes through the steps if you have it, then giving information of where to get help.’ (C)

When asked if there were any cultural improvements they wished to see in a mental health website, participants identified that the translation of the content into their ethnic language would be helpful. Interestingly, one participant mentioned the issue of cost.

‘... I know that there’s a cultural barrier especially with Pacific people and understanding health knowledge...I have a feeling it works...there’s still older Pacific people that still don’t have good English...I know it’s worked because my dad does work with the smoking thing and he’s helped people stop smoking with his resources and stuff and part of those resources are Pacific specific pamphlets.’ (D)
It's hard because I don't speak any language, I don't know personally if it would help. I think on behalf of like my grandparents it would be helpful but then that generation isn't good with computers anyway...I don't think there's any harm if it doesn't cost much. I think if you were to translate it, maybe something more hands on like a pamphlet or something. ' (A)

'Maybe if it offered it in maybe languages, so maybe in the option of their preferred language that they understand.' (B)

Participants had mixed reactions about which method they preferred to seek health information. Some preferred to see a doctor because of a lack of confidence with searching and the trustworthiness of online health information. Another preferred searching the Internet whilst one was not too sure because they had not practised searching health information online enough times to comment in greater depth.

'For me, I would just prefer to see a doctor. I think it's just coz I don't really know how to get health information online. I don't really do that, so I don't know...and confidentiality. I mean, yeah, you probably get that with a website as well but like if I was to find out more stuff, the doctor could, you know, help with that. The websites might have it but with a doctor you can ask whatever questions you want...and with that you can get confidentiality around your concerns and questions.' (D)

'I haven't really used it that extensively to answer that.' (A)

'The Internet. I think because it's quite a sensitive topic that it's not personal – I don't have to speak directly to someone, coz I still feel self-conscious with a sensitive topic. The Internet is my preferred choice but I think to get something to be more sure, I'd probably go to see my doctor.' (A)

When asked if they would search for online health information for another person other than themselves, the answer became dependent on the situation and context.

'I could look up [health information] like here are the symptoms for...but if I wanted a serious and in-depth knowledge about the issue...then I'd actually want to ask a doctor face to face. It would be good.' (D)

'It depends what it is, if it's something common I'd probably go to the doctor but if it isn't I'd definitely go online and find some information...I would probably look at service information...depending on what it is, I'd probably look at the condition first and how bad it is, and then service and treatment’ – common meant that cell needs to be checked out, go to the doctor, if someone said that's a cancerous looking cell I'd look into health information.' (A)

Although participants were mindful of using online health information as a source, it appears that because mental health is regarded as a serious health issue amongst Pacific Islanders, as opposed to a rash or infection, participant responses indicated that stigma issues are still present, possibly detracting from their usage of online health information. One participant noted that mental health issues were kept within a tight circle of family members. Further to this, she felt that being able to look at information online meant it held her responsible to look after the mentally ill family member. In other words it would be her burdening responsibility because she would then know where to source contacts, help and information. As a result family members may be heavily dependent on her to oversee the well-being of the mentally ill family member.

'I don't want to take on the responsibility [of looking after a mentally ill family member] if I find out the information online because where health is concerned, it becomes a family thing. I'd rather take them to the doctor’s with other family members so we can all talk, understand and ask questions, rather than me saying I found this, this is who we see, where we go...' (A)

When participants were asked about the advert used to promote the usage of the depression.org website, there was a general sense that the advert was not relatable to them. This was because Sir John Kirwan was not necessarily a figure they were able to relate to easily even though they knew who he was. With this in mind, these Pacific participants did not feel they needed to investigate or look into the website.

'I just feel like that ad was made for another group of people. I shouldn’t be listening to it.' (D)

'Oh, so like turn it off or turn the channel!' (B)

'Yeah or like go get a cup of tea or something [laughs] If there was a Pacific person standing there instead of him, I'd probably feel more comfortable to check it out coz I know we're the same people – they've probably experienced the same thing from being in the same situation and environment.' (D)

'I mean you see it [John Kirwan advert] but it's not really anything people take seriously if you're Pacific...but if I saw an Islander guy, I think so.' (C)

'Yeah probably, I never thought of it that way, I think it's enticing because it's John Kirwan, a famous face, but maybe. If it was someone like David Tua I think I would definitely be more interested to go in [and look at website] because it would be more relatable.' (A)

'Maybe the ad could be maybe with a PI [Pacific Islander]. It could not be just the individual but the family. You know how you either have two people or someone on their own...maybe advertise the help and how they went seeking for help...Maybe that's where PIs will think 'oh that's what we need to do ...maybe that could be the way to target the PIs because PIs are more family orientated. The story might help other families think 'well they did it, maybe we should'.' (A)

The above suggests that social marketing through television advertisements is being picked up by Pacific Islanders. These participants were able to identify that they were familiar with the advert, what he talks about and the website address he mentions. Despite this, participants felt they could not relate enough to investigate the website. Thus it appears that the underutilization of online health information by Pacific Islanders is also influenced by promotional material lacking cultural relativity. This means that there may be few aspects in social promotion surrounding online mental health information that Pacific Islanders can familiarise themselves with to investigate online health information websites.

4 Discussion

The purpose of this research was to examine how the attitudes and beliefs of Pacific people impact on their
propensity to use online mental health service [www.depression.org](http://www.depression.org). The findings from the questionnaire and paired interviews with four participants show that although people own computers and use them at work or as students they don’t seek mental health information or services online unless certain cultural considerations are integrated into the online services/information. The four female participants indicated that they did not relate to John Kirwan. This could be because he was male, older than they were, not a Pacific islander, and fronts the website alone (outside of the context of a family). One suggested that perhaps David Tua (a wrestler of Pacific island origin) could replace John Kirwan.

Trustworthiness of online information was an issue, especially considering the different views on what that could mean. Even if the trustworthiness were guaranteed, there were some disadvantages to seeking information online. If a person doesn’t understand what is written or said online they don’t have the same ability to pursue an understanding if they were talking to a doctor. It appears that education is an influential factor in the limited use of online mental health information amongst the New Zealand Pacific population.

Reasons for participants preferring not to use online mental health information were because they were unsure of how to identify trustworthy health information and how to search properly for health information. This could be because English is a second language for many Pacific people. Literacy problems further act as an explanation of why Pacific people prefer not to use online mental health information. The researcher notes that although each participant received a secondary education or higher, participants still identified they were unsure of how to search for mental health information or what to consider ‘trustworthy’. Although they gave good indication of what a trustworthy website is and were well educated, participants remained unable to search for health information online confidently. It could be noted that greater education around the usage of the Internet and computers is required. Education still remains a barrier for the uptake of online mental health information by Pacific people (Britannica, 2008).

Stigma and discrimination influence desire and comfort levels in seeking out information—taboo, low awareness, uncomfortable talking about people with mental illness, and negative connotations when referring to mental illness, e.g. ‘devil’s illness’. This may be due to the high incidence of mental illness among Pacific people relative to other cultures in New Zealand (Suualii-Suani et al., 2009).

Those who were not ‘New Zealand born’ would struggle with English and may benefit from being able to access online information in their own language. Getting information face to face from a doctor means that a family can do it together rather than an individual with Internet skills doing it on behalf of a family. The Pacific culture is socio-centric and relies on families working together for health improvement and maintenance, as evidenced in the different models of care (Suualii-Suani et al., 2009).

This interaction cannot happen with online mental health information. When using online health information especially regarding mental health, patients or users may be unable to get direct answers to specific questions and or understand the information provided. In either case the patient or user is forced to open further screens and other websites to find out bits of information to answer their question. This process alone may discourage the user from searching for health information and more specifically mental health information online.

Looking up information for a family member is associated with adopting responsibility for that family member.

**Limitations of the study**

This was a very small study - five key informants talking about building and delivering a e-mental health tools, i.e. [www.depression.org.nz](http://www.depression.org.nz), and four Pacific island people to answer a questionnaire and discuss their perspective of the website. The Pacific population within New Zealand accounts for 6.9% of the general population (Paterson et al., 2008). Of this 6.9%, the participation of four individuals does no justice in representing this population adequately. Furthermore, the participants are of ethnicities that fall into the Polynesian category of the greater Pacific. It is not known whether these answers of attitudes, beliefs and propensity are also reflective of the Micronesian and Melanesian ethnicities present in New Zealand. Each region has its similarities yet within each region the ethnic groups must be considered unique and independent of each other (Lal and Fortune, 2000, Thomas et al., 1989). Hence to help build a rich picture surrounding the usage (or lack of usage) of online mental health information for Pacific Islanders in New Zealand a greater number of participants incorporating the diversity of the Pacific is required.

The gender and age brackets of the women may have provided a biased outlook. Where questions about the advertisement, a promotional tool for social marketing, were asked, these were met with answers along the lines of ‘not being able to relate to it’. Participants were all female. Because the promotion surrounding online mental health information involved a male sportsperson this may have provided an instant issue of not being able to relate. Furthermore, more of the participants (three out of four) fell within the youngest age bracket of 24 to 30 years of age. Again, these participants may not have been able to relate to the promotion because of an elder figure delivering the message of using online mental health information. This visual presentation may not have appealed to the participants both because he was older than the participants interviewed and because John Kirwan was male.

**5 Conclusion**

In conclusion, the exploratory study resulted in some interesting findings. Although the researcher notes that the participant number was small, these findings call for further investigation. Such findings include:

- Relevance to the Pacific people in social marketing strategies surrounding online health information is necessary for increased uptake. Unless Pacific people are able to relate or align themselves with both the message and visual presentation, it is unlikely they will investigate online mental health websites for information.
Regardless of the level of education, gaps still exist around the usage of the Internet and computers. This further inhibits the use of online mental health information. This is due to confidence issues to explore the Internet and what is considered valid health information.

Mental health remains a sensitive topic amongst Pacific people, involving stigma and shame (Mak et al., 2007, Vaughan and Hansen, 2004). Although participants are aware of the discreetness and privacy the Internet provides when seeking mental health information, a doctor is still preferred as the source of health information. Reasons can relate to the above bullet point and the comfort found in face-to-face interaction regarding this sensitive health issue.

Further research is needed with a larger number of participants for replicate studies and the need to capture as many ethnic groups within the three regions of Polynesia, Melanesia and Micronesia to gain a more reflective stance on the issue of online mental health information usage for Pacific people.

6 References
Vaughan, G. & Hansen, C. 2004. ‘Like Minds, Like Mine’: a New Zealand project to counter the stigma and discrimination associated with mental illness. Australasian Psychiatry, 12, 113-117.
Health intelligence: Discovering the process model using process mining by constructing Start-to-End patient journeys

LUA PERIMAL-LEWIS¹, DENISE DE VRIES¹, CAMPBELL H THOMPSON²

¹ School of Computer Science, Engineering and Mathematics
Flinders University of South Australia

² Medicine
The University of Adelaide, South Australia

lu.perimal-lewis@flinders.edu.au, denise.devries@flinders.edu.au, campbell.thompson@adelaide.edu.au

Abstract

Australian Public Hospitals are continually engaged in various process improvement activities to improve patient care and to improve hospital efficiency as the demand for service intensifies. As a consequence there are many initiatives within the health sector focusing on gaining insight into the underlying health processes which are assessed for compliance with specified Key Performance Indicators (KPIs). Process Mining is classified as a Business Intelligence (BI) tool. The aim of process mining activities is to gain insight into the underlying process or processes. The fundamental element needed for process mining is a historical event log of a process. Generally, these event logs are easily sourced from Process Aware Information Systems (PAIS). Simulation is widely used by hospitals as a tool to study the complex hospital setting and for prediction. Generally, simulation models are constructed by ‘hand’. This paper presents a novel way of deriving event logs for health data in the absence of PAIS. The constructed event log is then used as an input for process mining activities taking advantage of existing process mining algorithms aiding the discovery of knowledge of the underlying processes which leads to Health Intelligence (HI). One such output of process mining activity, presented in this paper, is the discovery of process model for simulation using the derived event log as an input for process mining by constructing start-to-end patient journey. The study was undertaken using data from Flinders Medical Centre to gain insight into patient journeys from the point of admission to the Emergency Department (ED) until the patient is discharged from the hospital. This paper presents a historical event log of the patient journey from the point of admission to the Emergency Department (ED) until the patient is discharged from the hospital.

Keywords: patient journey, process mining, simulation model, event logs, hospital key performance indicators, emergency department (ED), general medicine (GM), inliers, outliers.

Introduction

The demand on Australian Public Hospitals continues to intensify as the life expectancy of the overall Australian population increases especially the rise in residents who are 65 years old or older. The trend in South Australia as reported by Banham et al. (2011) from a study undertaken from 1999 to 2008 showed that both total life and healthy life expectancy increased from 2.0 years among males and 1.5 years among females to 1.4 years among males and 1.2 years among females respectively. According to the Australian Institute of Health and Welfare (2011-2012) reported that there was an increase of 4.3% on average each year between 2007–08 and 2011–12 in Emergency Department (ED) presentations accounting for over 6.5 million presentations. The department also reported that the overall proportion of ‘seen on time’ were 54% in Northern Territory and 76% in New South Wales and South Australia. 50% of patients received their treatment within 21 minutes and 90% received their treatment within 108 minutes of presentation to the ED. 28% of the ED patients were admitted as hospital inpatients to continue care after the completion of ED treatment.

The data presented in the paragraph above shows the imminent increase in the demand for services at Australian Public Hospitals. Proportion of patients ‘seen on time’ is one of the ED Key Performance Indicators (KPIs) used Australia wide as a measurement of ED efficiency. The increased demand for services together with the pressure of progressing patients within a set timeframe in an attempt to satisfy certain KPI requirements often contributes to patients being streamed to wards that are not equipped with all the required facilities to treat the patient’s conditions. Patients who stay outside of their homewards are referred as outliers.

EDs around Australia are often brought into the limelight by the media as being affected by access block. A patient is considered to have experienced access block if the patient has waited in the ED for 8 hours or more waiting for an inpatient bed and General Medicine (GM) patients are more likely to experience access block (Perimal-Lewis et al., 2013). O’Connell et al. (2008) assert that ED congestions are intensified by regular
failure to manage processes involved in progressing patients through the hospital and highlight that the lack of shared understanding among staff, patients and carers of the probable patient path as one of the contributing factors amongst others. The authors call for better inpatient management for a better flow which will in turn ease issues faced within ED.

Hospitals are continually trying to improve their processes in order to cope with the increase in the demand for service and to increase the efficiency of delivery of care. Hospitals need to go beyond conventional aggregate information produced which is collected as part of performance reporting for the in-depth knowledge needed for process improvement activities. In previous work in this area Perimal-Lewis et al. (2012) stated that Clinical Process Re-engineering (CPR) is considered similar to Business Process Re-engineering (BPR) where both activities focus on continuous improvement to core business or clinical processes. The authors also regarded patient journeys as the core business process for hospitals and differentiated it as being patient-centred and service-oriented rather than business-oriented. Ben-Tovim et al. (2008) stated that clinical process redesign aims to harmonise the poorly coordinated patient journeys as patients move across multiple departments taking a holistic approach by looking at a wider area during the redesign process.

The performance of public hospitals is compared and judged publically by certain KPIs. Conformance to KPIs may soon become essential because of competitive government funding. KPI reporting is generally presented as an aggregate figure, such as ‘20% of patients are seen on time’. Often a poor KPI reported on core business areas would initiate a closer look at the underlying processes to investigate and redesign the processes in order to improve performance.

Processes in hospitals are complex. Process mining aims to gain insight into a process from carrying out detailed analysis using historical event data pertaining to that process. Generally, these event logs are easily sourced from Process Aware Information Systems (PAIS). Process mining is a Business Intelligence (BI) tool which aims to improve the operational business processes by amalgamating the knowledge from information technology and management science as defined by Van der Aalst (2011) who is also the pioneer of this field. As such applying process mining in health will contribute to in depth analytical knowledge contributing to Health Intelligence (HI). Unlike many mainstream BI and data mining tools which are data-centric, process mining is process-centric aimed to gain insight into the processes the data refer to and the focus is not on fancy-looking dashboards rather a deeper analysis of the data (Van der Aalst, 2011).

This paper outlines how unstructured event data were processed to derive the event logs needed as an input for process mining in the absence of PAIS. Using the processed event data, process mining was then applied for an evidence-based process model discovery of patient journeys from start to end at Flinders Medical Centre (FMC). The discovered process model is the base for a simulation model. The discovered process models could also be used for performance analysis and verification. The knowledge discovered using process mining techniques could be used as a basis for process optimisation.

The remaining section of the paper is structured as follows:

Section 2 gives a brief background to the conventional ways of modelling used in healthcare. Section 3 gives information on the study setting and the origin of data for the case study used in this paper. Section 4 describes the methodology used to derive the event log for process mining in the absence of PAIS. Section 5 describes only pertinent results related to the argument of this paper where the final output is the petri net model for simulation project. Section 6 is the discussion and finally the paper is concluded in Section 7 with the conclusion and future work.

2 Background

Traditionally health care modelling has been done using various mathematical modelling techniques focused on forecasting and predicting in order to improve health care performances (Perimal-Lewis et al., 2012). New approaches which use a combination of techniques to complement the strengths and weaknesses of the one technique are quickly emerging. One such research was carried out by Cegielski et al. (2007) who proposed combining Data Mining techniques and discrete event simulation for identifying bottlenecks in the patient flow between ED and a hospital ward by providing insight into the complex relationship between patient urgency, treatment and disposal and the occurrence of queues for treatment.

Simulation is widely used in health care as a basis to understand processes and for prediction. There are many simulation projects which are discontinued after implementation as these models fail to improve the underlying processes. Process models are the core component of any simulation projects; therefore it is essential for the models used for simulation to give a close reflection of reality. Creating a simulation model to depict reality by hand is a challenging task especially in a complex healthcare environment where the system is prone to numerous process variations. The conventional ways of creating process models by hand, ignoring event data are error prone which can lead to wrong conclusions. Using process mining solves this issue as models are extracted from events that have already taken place giving a close reflection of reality (Van der Aalst, 2011).

Mans et al. (2008) used process mining techniques to better understand different clinical pathways taken by diverse groups of patients and used these techniques to identify bottlenecks. Rebuge and Ferreira (2011) concluded that although process mining techniques have been proven in some instances as being successful in mining health data, there is still room for improvement to identify the right algorithm to handle noise in the data, complexity of data and the ad hoc nature of health data.

The proposition made in this paper is to use process mining techniques for the discovery of process model/s from historical event data which form the foundation for a simulation model. This model would give a closer reflection of reality. This paper introduces a novel way of
deriving event logs for process mining used to discover a process model that is validated by the domain experts as an authentic representation of the patient journey from start-to-end. The discovered process model could then be used as an input model for simulation projects.

3 Study Setting & Data
The analysis was undertaken on inpatient records for patients admitted and discharged by the General Medicine (GM) service at Flinders Medical Centre (FMC) and one specialty unit which is the Cardiology Unit. FMC is a public teaching hospital in South Australia and it attends to approximately 62,000 admitted inpatients per annum. The two largest medical inpatient specialties are GM and Cardiology. The typical patient admitted to each of these units is very different in terms of their age, complexities of disease and diagnosis. The GM service looks after a wide variety of diagnoses. The GM service controls about 100 inpatient beds out of about 500 beds in FMC as a whole. Cardiology is a specialty unit looking after a limited number of specific diagnoses that treats the highest number of patients compared to other specialty units. The analysis was carried out on inpatient records of the GM service; that is, on those patients whose inpatient care had been allocated to a GM team and on inpatient records of the Cardiology service; that is, on those patients whose inpatient care had been allocated to a Cardiology team. The wards that were ‘home-wards’ for this service were clearly defined. A ‘home-ward’ is a ward that is equipped with the appropriate medical team and specialised equipment to treat the patient’s primary disease. Patients who were not allocated a ‘home-ward’ of the medical units responsible for their care were defined as being an outlier and stayed in an outlier ward. A similar concept is applied to any other specialty units.

The Patient Journey Database from FMC contains information on inpatients or officially admitted patients only and records detailed information on the journey or movements of a patient from the time of admission to the time of discharge. An individual patient could have multiple admissions at different points in time and each admission will be allocated with a unique journey number that remains the same until discharge. Each movement of the patient from one ward to another ward is recorded with a timestamp, so at any point the “start time” in a ward and the “end time” in a ward are known together with the name of the ward. Each ward occupied by a patient is appropriately marked to reflect whether the ward occupied was an inlier or an outlier ward. A patient admitted to an inlier ward is admitted to their ‘home-ward’. The timestamp for Admission is the combination of the “Date” field and the “Admission Time” field. The timestamp for Discharge is the combination of “Date” field and “Discharge Time” field. The timestamp is a derived field. The individual patients are not identifiable at any point.

Ethics approval for the use of data was granted by the Southern Adelaide Health Service / Flinders University Human Research Ethics Committee.

4 Methodology
Process mining activities can be categorised into three different perspectives, which are the process perspective, the organisational perspective and the case perspective. (Weijters et al., 2006). The process mining category presented in this paper is the process perspective. The goal of a process perspective is to focus on the control flow or the ordering of activities with the intention of discovering all possible paths (Weijters et al., 2006). The key to producing a good simulation model is to first understand the model with all possible variations which cater for all scenarios to produce a model that will be as close to reality as possible. A model accounting for all variations would help the domain experts perceive the entire process as it has taken place. Once the entire model is produced, it can then be deduced to represent the major behaviour of the system.

For a complex process such as the healthcare environment formulating a process model close to reality is far from trivial, therefore the use of historical data to derive a process model for simulation is advocated. This section describes the methodology used to develop a petri net based simulation model for the GM inpatients and for the Cardiology specialty unit’s inpatients. The main software used for Process Mining is ProM (The Process Mining Group, 2010). ProM is open-source specialised process mining software. The constructed event log was preprocessed into the MXML format required as input to ProM. This conversion was done using the Disco software package (Fluxicon, 2012). Other ancillary software used was MS Excel and MS Access which was also necessary for the pre-processing of the event log. In section 4.1, the hospital admission process as explained by the domain experts is described to set the context of the area and process being investigated. The knowledge of the area and the underlying process is the foundation for constructing the event log which is described in section 4.2. Then in section 4.4 after the required event log is derived, the application of one of the process mining algorithms within ProM to derive the process model for simulation is described.

4.1 FMC’s admission process
FMC offers both inpatient and outpatient services. Outpatients are seen during business hours at the outpatient clinics and sometimes these patients might be admitted as an inpatient. The inpatients could be categorised into two streams: those who enter the hospital as an emergency admission and those whose admissions are pre-planned for the elective surgery stream. Both streams of inpatients affect the hospital occupancy.

The time patients spend in the ED can be categorised into three distinct processes. The phases related to these processes are categorised as Waiting to be seen (FMC-WTS), Assessment time (FMC-RT) and Boarding (FMC-Boarding). The time patients spend at FMC-WTS and overall ED waiting time (= FMC WTS + FMC RT + FMC-Boarding) is measured and reported by the hospital as these times are part of the hospital’s KPI. All the three phases of time take place within the ED.

The flow chart in Figure 4.1, as illustrated by the domain experts, is the reflection of the three ED phases.
portraying how patients flow through the ED and either end up as an inpatient or are discharged from the ED.

An understanding of the processes and how patients flow through the ED and the KPIs surrounding these processes was a starting point for the process mining activities. This knowledge is essential in order to identify the data set and fields required to construct the inpatient journeys discussed in the next section. The flow-chart in Figure 4.1 as depicted by domain experts is used to verify the discovered process model/s and could also be used for conformance checking.

Figure 4.1: GM patient journey flowchart

4.2 Process Mining – Feature Extraction

In order to gain a holistic view of the patient journey process, it was necessary to derive an event log containing the required fields needed to discover the operational process model of patient journeys. The process of deriving the operational event log is far from a trivial exercise. The sensitive nature of health data and the ethics laws surrounding the use of health data meant that access to the information systems were not practical therefore the historical raw data supplied were sourced from various databases by the hospital. The raw files supplied were in both comma-separated values and tab-separated flat files. The records did not conform to any database structure. The required features from these flat files were extracted to form the patient journey event log used for process mining. Figure 4.2 shows a snippet of the patient journey tab-separated flat file supplied.

Figure 4.2: Patient journey tab-separated flat file

The patient journey dataset only contains data on an inpatient’s ward movement from the time of admission to the time of discharge, therefore only data on officially admitted patients are recorded. The patient journey data supplied had to be transformed and merged with another set of raw data which contained patient data whilst in the ED. The ED dataset contained data on activities that took place on a patient’s journey in the ED up until the patient was officially admitted to the hospital. The snippet of the ED data comma-separated flat file is shown in Figure 4.3.

Figure 4.3: ED data comma-separated flat file

The raw data from both flat files had to be transformed in the first instance and then merged to form the patient journey event log. MS Access 2010 was used to merge both datasets into a single database. Upon merging the dataset, the semantics of each field was established. Each field had to be converted to its appropriate data type. The major challenge for this process is the conversion of the date/time fields. In the patient journey dataset both ‘time1’ and ‘time2’ fields are recorded as minutes past midnight. The timestamp for the event log was derived by concatenating the ‘date’ field and ‘time1’ field to form a new field called ‘DateIn’ field. The ‘DateOut’ field was formed by concatenating ‘date’ field and ‘time2’ field. Prior to concatenating the fields, the ‘date’ field had to be converted into a Date/Time data type with a 24 hour notation. For example:

Format ([time1], "General Number")

After calculating a timestamp from ‘time1’ and ‘time2’, these values were concatenated with the date value as below:

DateIn: [date] & " " & [time1] \rightarrow Timestamp admission.

DateOut: [date] & " " & [time2] \rightarrow Timestamp of discharge or timestamp at the end of 24 hours.

The concatenated fields were then converted into date/time data type. Accurately deriving the timestamp is essential for an accurate discovery of process model using the process mining tool, ProM. Each patient is uniquely identified by the Patient Unit Record Number ‘URN’ field in both datasets. The data type for this field had to be converted to the same data type in order to merge the records. In the patient journey dataset, each admission is identified with a unique ‘journey_id’. The ‘journey_id’ is unique for a particular admission and stays the same until discharge. Multiple admissions by the same patient will have multiple ‘journey_id’ s. As a result, when merging the two datasets, two fields had to be used as identifying keys. The first key is the unique ‘URN’ in both datasets.
The second key is the derived ‘DateIn’ field from the patient journey dataset and the converted ‘Outcome Date’ field from the ED dataset. The ‘Outcome Date’ is the date when the patient is either admitted as an inpatient or discharged from the ED. If the patient is admitted, the patient’s record will be recorded in the patient journey dataset.

The ED dataset contained three date and time fields: ‘Triage Date’, ‘Date Time Seen’ and ‘Outcome Date’. All fields had to be converted to the date/time data type. The ‘Triage Date’ is the timestamp when a patient enters the ED and is triaged according to treatment priority. The ‘Date Time Seen’ field is when the patient is seen by a doctor. All three timestamps are essential components in depicting the process that takes place in the ED. The ED time is broken into three phases as described in Section 4.1: elapsed time between triage and when the patient is seen by a doctor (FMC-WTS), elapsed time from being seen by the doctor until a decision is made to admit the patient which is the overall assessment time (FMC-RT), and the time spent in the ED waiting for an inpatient bed after the decision to admit is made which is the boarding time (FMC-Boarding). All three components: FMC-WTS, FMC-RT and FMC-Boarding have been derived from the raw ED dataset. The snippet of the derived event log is shown in Figure 4.4. For each journey, the three components of ED time are derived from the ED dataset and the ward movement data are derived from the patient journey dataset. This extracted event log now contains information on the inpatient journey process from start to end.

The cases in the patient journey process are the individual journeys. A patient is identified using the ‘URN’ field which is one of the attributes of the case. A patient could have multiple journeys or cases (multiple admissions). As this particular event log is derived to construct the inpatient journey from ED to discharge, the journey starts from the point of triage. Within the ED there are three distinct events which relate to distinct ED processes as described in the previous section: FMC-WTS, FMC-RT, and FMC-Boarding. These ED processes also relate to the KPIs reported by the ED hence the availability of event data to produce the aggregate statistical information. Once the ED processes are completed the patients are then moved to either an appropriate ward or to any available ward. In the context of this event log an activity relates to the patient moving within the predefined structured process (e.g. progressing from one activity/phase to another within the ED followed by the process of ward movement). Each event or activity is uniquely identifiable by the ‘EventID’ field. Each ward within the hospital has a predefined activity with regard to the patient’s care. Generally, an inpatient changes ward if he/she is in an outlier ward. Each event or activity (e.g. ward movement) is recorded with a timestamp (the ‘DateIn’ and ‘DateOut’) field. These fields are some of the many possible attributes that could be derived for an event. The timestamp of the event reflects the order of patient movement.

4.4 Process mining – Heuristics Miner - algorithm

It is apparent from the GM patient journey flowchart (Figure 4.1) that the patient journey flow from the ED to various wards outside of the ED is a structured process. Based on this knowledge, the patient journey flow process was characterised as a “Lasagne Process”. In a “Lasagne Process” most cases are handled in a structured and pre-arranged manner (Van der Aalst, 2011). For example, certain pre-conditions have to be satisfied before the patient can move to the next activity/phase. The process mining algorithm within ProM, namely the Heuristics Miner was used to discover the control flow of the patient journey process from admission to discharge. It is important to note that the order of activities within each case (each admission to discharge) is important as this information is used to calculate the order of activities (the order of ward movement). In other words, the algorithm relies heavily on the timestamps (DateIn and DateOut). The Heuristics Miner-algorithm is briefly described to set the context for the models presented in Section 5. The finer details of the workings on this algorithm are addressed in “Process Mining with the Heuristics Miner-algorithm” (Weijters et al., 2006).

The control flow process model is constructed by analysing for causal dependency. In this context the event log is analysed to see if a patient staying in a particular ward always moves to another particular ward. If this movement frequently occurs then there is a causal dependency between these two wards. As described by Weijters et al. (2006), the dependency graph is constructed by:

<table>
<thead>
<tr>
<th>Journey No.</th>
<th>Timestamp</th>
<th>Priority</th>
<th>DateIn</th>
<th>DateOut</th>
<th>Ward/Def</th>
<th>Unit/Def/End</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>27/01/2004 10:30</td>
<td>1</td>
<td>5</td>
<td>T4A</td>
<td>FMC-WTS</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>27/01/2004 14:10</td>
<td>1</td>
<td>5</td>
<td>T4W</td>
<td>FMC-R</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>27/01/2004 10:30</td>
<td>1</td>
<td>5</td>
<td>T4A</td>
<td>FMC-WTS</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>27/01/2004 10:30</td>
<td>1</td>
<td>5</td>
<td>T4A</td>
<td>FMC-WTS</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>27/01/2004 10:30</td>
<td>1</td>
<td>5</td>
<td>T4A</td>
<td>FMC-WTS</td>
<td></td>
</tr>
</tbody>
</table>

Figure 4.4: The derived event log

Each tuple contains information about an event. The next section will discuss this derived event log and how it relates to the process mining methodology and concepts as discussed by Van der Aalst (2011).

Once a systematic method is established to generate the event log of a process of interest, attributes for cases and activities could be easily extracted for further analysis of cases. An event log contains millions of records, therefore for meaningful analysis it will be necessary to filter the event log according to a specific scope or boundary to produce models that are interpretable.

4.3 Process information from event log

Van der Aalst (2011) makes the following assumptions about event logs: a process consists of cases, a case consists of events such that each event relates to precisely one case, events within a case are ordered, events can have attributes. The derived event log of the patient journey process conforms to this assumption.
Process mining is an iterative technique. The main challenge was establishing boundaries for the underlying processes surrounding important hospital KPIs. Identification of these KPIs helped the domain experts with identifying the relevant data set needed to be extracted from various systems used by the hospital. Using this data a specific event log was constructed as described in section 4.2.

Once the event log which is the most fundamental element in a process mining activity is available, this event log could be used for knowledge discovery applying various process mining algorithms available within ProM. The result presented here is specific and limited to the discovery of a simulation process model and corresponding analysis. The same event log could be used with various other process mining algorithms and analysis features within ProM which is not presented in this paper. Section 5.1 below presents the descriptive statistics in relation to the ED and the processes under investigation.

5 Results

Process mining is an iterative technique. The main challenge was establishing boundaries for the underlying processes surrounding important hospital KPIs. Identification of these KPIs helped the domain experts with identifying the relevant data set needed to be extracted from various systems used by the hospital. Using this data a specific event log was constructed as described in section 4.2.

Once the event log which is the most fundamental element in a process mining activity is available, this event log could be used for knowledge discovery applying various process mining algorithms available within ProM. The result presented here is specific and limited to the discovery of a simulation process model and corresponding analysis. The same event log could be used with various other process mining algorithms and analysis features within ProM which is not presented in this paper. Section 5.1 below presents the descriptive statistics in relation to the ED and the processes under investigation.

5.1 Descriptive Statistics

Descriptive statistical analysis of the overall FMC’s patient data shows predictable patterns. Various statistical analyses are already being carried out to improve the efficiency and quality of patient care in general. One such previous work relating to ‘Quality of Care’ received by inlier and outlier patients has been addressed (Perimal-Lewis, 2013). Figure 5.1 and 5.2 show the trend in average waiting time and average patient count at the ED during triage.

Average waiting time is the time patient spent in the FMC-WTS phase which shows an obvious association with the patient count in the ED at the time of triage. For example on Mondays, when there are more patients in the ED, the average waiting time also increases.

Figure 5.1: Trend in average waiting time (FMC-WTS)

Average patient count at triage time

Based on the data captured for the various ED KPI reporting, FMC is already undertaking sophisticated statistical analysis for different prediction models. However, FMC is still experiencing ED congestion and access block. Process mining could be used to complement the already mature statistical analysis to understand this process that extends beyond the aggregate statistical analysis. By undertaking process mining, it is possible to dive deeper into the processes underlying the inpatient journey or patient flow from admission to discharge. The next section discusses the heuristic models for GM patient and the Cardiology specialty unit. Similar models could be built for other units as required.

5.2 Control flow perspective – heuristic models

As the models discovered are the base for simulation model/s, it was essential to choose a timeframe within the data set that reflected the processes within FMC where there was stability. Deriving this timeframe was done with close consultation with the domain experts. The timeframe used for these models was between 01/01/2007 and 31/12/2009.

Figure 5.3 shows the heuristic model for the cardiology specialty unit. The model is a less complicated model compared to the GM patients’ model which will be discussed next. For the purpose of simplicity the model is a reflection of cardiology patient flow from 01/01/2007 – 21/12/2007 only, with further filtering of records to show patients that received 100% of their care from the Cardiology team. This means that these patients would have received their entire care from the same team of doctors. This is a rare but good occurrence as fewer unit (team of doctors providing care) changes are better for
the patients as the patients would receive undisrupted care. The model is verified by the domain experts to be a correct reflection of patient journey for the Cardiology unit. The weightings next to the arcs between the wards indicate whether there is a strong or weak dependency between the wards as described in section 4.4. As reflected in Figure 5.3 the dependency relationship is not too strong as firstly only a small subset of patients are modelled. Secondly, the possibility of such a transition where a patient’s care from ED to discharge stays with the same team of doctors is often rare.

Figure 5.3: Cardiology patient journey

The next model shown in Figure 5.4 is the first model discovered for the overall GM inpatients’ journey. As stated before the model building exercise was an iterative process. The first model derived is as represented in Figure 5.4 which had high variation. Although the processes are well defined, there were high variances in the event log. This was acknowledged and explained by the domain experts as the nature of GM patients. There will always be patients presenting to the hospital with a unique characteristic that would require the patient to follow a unique path. The discovery of the complicated model confirms the perception of the complex nature of GM patient journeys. Revealing the complexity of the GM patient journey was an important exercise however using a model with such high variation will not be beneficial in deriving a simulation process model. The GM patients’ journeys will always have high variation. Therefore, in this situation a model that portrays the majority of the patient journeys will be a better model. Further analysis of this model using ‘Performance Sequence Diagram’ within ProM revealed that there were over 2000 path patterns and many one-off paths. One-off paths do not show the main behaviour of the system. The domain experts verified the model and confirmed the validity of the variations. However modelling the paths that reflected the common behaviour of the system was deemed important in order to identify paths or patterns with high throughput and paths that could contribute to bottlenecks in the system. With this notion, a second model for the GM inpatients was developed as shown in Figure 5.5 and for better readability a small section is of the model is shown in Figure 5.6

Figure 5.4: Complexity of first patient journey process model for GM patients

Figure 5.5: Complexity of the second patient journey process model for GM patients

The model in Figure 5.5 shows the second heuristic model for the GM inpatients similar to the previous model. However this is a less complex model. The model presented is the final and most representative model. The model is based on GM inpatients where the sequence of activities (the path) is shared by at least 10 cases. This means, journeys with a full path from start to end that appeared less than 10 times were filtered out in order to produce a model that is interpretable in a complex setting such as the hospital. The model accounted for 75% of the GM inpatients and was more interpretable. The number of patterns was now 113 patterns rather than 2000 patterns or more in the previous model. The model was verified to be a good reflection of the GM patient journeys by the domain experts. Similarly, a model representing 80% of the population could be derived and validated by the domain experts if necessary.
The second heuristic model in Figure 5.5 and the snippet of the model in Figure 5.6 are also close representations in conformance with the GM patient journey flowchart shown in Figure 4.1. This knowledge helps enforce the validity of the discovered model. As well as validating the process model as depicted by the domain experts, the discovered model also revealed other ward movement patterns which were not accounted for by the domain experts. The model also helped identify potential deviations in the process which could also be attributed to data entry errors. For example patients should not be moving back to FMC-WTS phase from FMC-RT phase as reflected in the models. However this only accounted for a very small percentage of patients.

Furthermore all the causal dependency values for the first and second GM patients’ models were more than 0.9 indicating that the dependency relationships between wards are strong, enforcing more confidence in a particular pathway as being a common feature of the GM patient journey. Other movements were verified correct by the domain experts to depict the GM patient journeys. Based on the discovered process model, it was also possible to verify the inlier and outlier wards where GM patients were admitted. Further analysis which is beyond the scope of this paper could be carried out to analyse the characteristics and outcome of patients following a path consisting of mainly outlier wards as opposed to paths consisting of mainly inlier wards.

The model also depicted wards with a high percentage of unit changes which were verified to be a correct reflection of wards where the care of a patient might be transferred to another team because either the patients were being wrongly diagnosed and hence "sorted" into that unit or the unit offered a higher acuity of care and a significant deterioration in patients’ condition often required a change in the team of doctors looking after those patients.

The second heuristic model in Figure 5.5 and the snippet of the model in Figure 5.6 are also close representations in conformance with the GM patient journey flowchart shown in Figure 4.1. This knowledge helps enforce the validity of the discovered model. As well as validating the process model as depicted by the domain experts, the discovered model also revealed other ward movement patterns which were not accounted for by the domain experts. The model also helped identify potential deviations in the process which could also be attributed to data entry errors. For example patients should not be moving back to FMC-WTS phase from FMC-RT phase as reflected in the models. However this only accounted for a very small percentage of patients.

Furthermore all the causal dependency values for the first and second GM patients’ models were more than 0.9 indicating that the dependency relationships between wards are strong, enforcing more confidence in a particular pathway as being a common feature of the GM patient journey. Other movements were verified correct by the domain experts to depict the GM patient journeys. Based on the discovered process model, it was also possible to verify the inlier and outlier wards where GM patients were admitted. Further analysis which is beyond the scope of this paper could be carried out to analyse the characteristics and outcome of patients following a path consisting of mainly outlier wards as opposed to paths consisting of mainly inlier wards.

The model also depicted wards with a high percentage of unit changes which were verified to be a correct reflection of wards where the care of a patient might be transferred to another team because either the patients were being wrongly diagnosed and hence "sorted" into that unit or the unit offered a higher acuity of care and a significant deterioration in patients’ condition often required a change in the team of doctors looking after those patients.

Finally, once the verification of the models is to the satisfaction of all concerned, these models were converted into a Petri Net model. Figure 5.7 is a snippet of a Petri Net model which was derived from the second patient journey model for GM patients shown in Figure 5.5. The Petri Net model can now be exported into a simulation tool such as the Coloured Petri Net (CPN) tool for simulation.

Figure 5.7: Snippet of Petri Net for GM patients

Coloured Petri Nets (CPNs) are a discrete-event modelling language for modelling systems where concurrency, communication and synchronisation plays a major role (Jensen et al., 2007). The CPN process model that will be used for simulation is a sound model derived from historical event logs and validated as a close reflection of reality by the domain experts. As a result, the patient journey process models discovered gives high confidence into the output of simulation exercise as these models are based of event data that has already taken place. Also the process model reflects the main behaviour of the system and reduces the chances of excluding certain activities by mistake when constructing these models by ‘hand’.

6 Discussion

It is important to be mindful of the scope and boundary for the data needed as otherwise the big data files available in health sector could pose not only technical difficulties requiring high end computer processing power but also could produce models that are not interpretable. Undefined scope for process mining in health care could lead to discontinuation of such projects. Therefore the collaboration with the domain experts should start at the very inception of the project and continue at every stage of the project.

Compared to inpatients of a single specialty unit, the inpatient characteristics of GM patients are complex and are non-deterministic. Therefore, studying and understanding the underlying processes of the GM patients although challenging will reveal insight to wider spectrum of behaviour of the patient journey process. Insight such as this from the domain experts is significant for successful application of process mining in healthcare settings. Choosing a diverse patient group and then focussing on single specialised unit with less diverse patient separately is an important strategy for the big hospital data.

A distinction is made to differentiate the characteristics of the event log and the characteristics of the process being mined. The event log used is characterised as unstructured or semi-structured, however the process being mined is characterised as being structured. This distinction is important in deciding the appropriate control flow algorithm to use within ProM. The process is structured because there are pre-defined
activities and criteria that take place under each phase following a structured process. For example patients under the FMC-WTS phase are treated according to the Australian Triage Scale (ATS) which is a measure of severity and in the order of arrival.

The simulation model discovered was for a specific group of inpatients. Similar models could be discovered for other groups of patients.

7 Conclusion & Future Work

Process mining in the healthcare domain is an extensive and time consuming exercise. For process mining activities to be successful the stakeholders involved need to perceive the advantages of using process mining for gaining health intelligence. Health intelligence is gained by diving deep into a process for knowledge discovery beyond what is offered by statistical analysis alone. The complex movement of patients show that patient journey analysis using a statistical approach combined with process mining techniques will give better insight into the intricacies of a complex healthcare system.

The next challenge is to work closely with the domain experts to identify key process improvement areas. For healthcare, these areas are normally areas where the performances are measured by KPIs. This will lead to the identification of appropriate data needed which will be used as an input for process mining activities. The identification of appropriate data and then the process of pre-processing the data to derive the event log is the most crucial and time consuming activity. In most healthcare settings within Australia the absence of Process Aware Information Systems (PAIS) means that a resourceful way of deriving these event logs is needed for process mining activities to be successful. This paper presented one such method for deriving event log in the absence of data warehouse or PAIS. In such instances, using the available data used for KPI reporting could be a starting point. Since similar KPIs are used in all public hospitals in Australia, this method is generalizable and similar approach could be used for other hospitals that would like to embrace process mining to dive deeper into their process as a starting point for process improvements.

At FMC, process mining offers added benefit to the already successful implementation of “lean thinking” and enhances the areas where a “lean thinking” approach alone is inadequate to investigate and reveal insights to access block and bottlenecks. Therefore constructing a process model for patient journeys from start-to-end as a base for simulation model derived from historical event log and validated by domain experts is advocated as a sound starting ground for future simulation projects.

The final start-to-end patient journey process models discovered as a base for simulation discussed in this paper could be further extended by including petri nets from other process mining perspectives such as the organisational perspective and the case perspective. The entire integrated model would give a holistic view of the process and therefore produce an all-encompassing input for simulation projects.

8 References


Abstract

Association Rule Mining (ARM) is a promising method to provide insights for better management of chronic diseases. However, ARM tends to give an overwhelming number of rules, leading to the long-standing problem of identifying the ‘interesting’ rules for knowledge discovery. Therefore, this paper proposes a hybrid clustering-ARM approach to gain insight into a population’s pattern of risk for a chronic disease related adverse event. Our current experiment is based on the Framingham Heart Study dataset and we focus on Myocardial Infarction (MI, ‘heart attack’) as the adverse event. Association rules indicative of MI are developed from training data and clustered based on commonality of cases satisfying the rule antecedents. Test cases are then assigned to the rule clusters to provide sets of at-risk patients sharing common MI risk factors. We demonstrate this approach for a range of clustering methods and cluster counts, illustrating some of the derived participant sets.

Keywords: Adverse event modelling, Association rule mining, Chronic disease management, Clustering.

1. Introduction

There is a great need to better understand patterns of chronic disease occurrence and progression in populations, particularly with an eye to potentially preventable events. Prediction of hospital readmission risk, for instance, is an area in need of better-performing models to target care transition interventions (e.g. support for patients after hospital discharge) and to calibrate the benchmarking of health service performance (Kansagara et al., 2011). Indeed, chronic diseases are associated with more than 50% of all potentially preventable hospital readmissions in Australia (AIHW, 2009).

In the present study we focus on the question of identifying relatively disjoint sets of patients at high risk of potentially preventable chronic disease events. The goal of identifying such sets is to provide novel descriptions of chronic disease patterns that may guide healthcare providers in formulating preventative responses. We use the freely-available anonymised (‘teaching’) Framingham Heart Study cohort data set (from the Biologic Specimen and Data Repository Information Coordinating Center, BioLINCC) for illustrating of our approach to this problem. We take Myocardial Infarction (MI, a ‘heart attack’) as our adverse event outcome as this is familiar, well-documented and significantly preventable by known interventions such as blood pressure control and smoking cessation (New Zealand Guidelines Group, 2012).

Traditional regression model analysis (including logistic regression and survival models) identifies significant risk factors across the entire population. While this is useful, the focus is on factors rather than characterisation of patient groups. Moreover, regression models are relatively clumsy for identification of combinations of risk factors (e.g. by manually introducing interaction terms). In contrast, Association Rule Mining (ARM) is more appropriate to explore the association of combinations of factors to groups of patients with high event risk. However, ARM tends to give an overwhelming number of rules and there is the long-standing problem of deciding an ‘interestingness’ measure (Tan and Srivastava, 2002) to filter the rules for knowledge discovery. In order to gain meaningful insight into patients with chronic diseases by means of ARM, we hereby propose applying a ‘Cluster-ARM integrated framework’ wherein the proliferation of association rules is managed by clustering on commonality of cases satisfying the rule antecedents. Groups of at-risk patients are then identified and characterised by their association to these rule clusters.

The following section of this paper presents some general definitions for understanding of our approach. In section 3, we introduce our approach in general, including data description and staple data pre-processing. That is followed by section 4 and 5.
involving a detailed explanation of the proposed Cluster-ARM integrated framework implementation and results on the illustrative data, with brief conclusions and future research in section 6.

2. Clustering and ARM

Clustering and ARM are data mining techniques for different purposes. Clustering explores groups of similar points according to some similarity metrics (Bradley, P. et al., 1998). ARM discovers frequent patterns and correlations within frequent patterns (Agrawal et al., 1993).

Classification Association Rule Mining (CARM) (Quinlan et al., 1993) is a Classification Rule Mining approach by means of ARM. CARM mines a set of Classification Association Rules (CARs) from a classified transaction database, where each CAR describes an implicative (although not necessarily causative) relationship between a set of data attributes and a pre-defined class.

3. Approach

In 1995, Riddle et al. proposed a generic knowledge discovery framework for high dimensional data mining. In this framework, they suggested three steps: 1) to develop a statistically valid process for selecting and ordering interesting rules; 2) to develop a summarization and visualization methodology to help experts examine rules; and 3) to develop a methodology for deriving data representations which best captures the objectives of process owners (Riddle et al., 1995). Later, Lent et al. applied the framework on an association clustering system for two-dimensional ARM specifically (Bannink et al., 2006).

In this research, we apply the same framework on generating, managing and evaluating multi-dimensional association rules on diverse outcomes of chronic diseases. Our current experiment is based on the Framingham Heart Study cohort dataset and we focus on MI as the outcome of chronic disease. The dataset is divided into training and testing parts. After data pre-processing, ARM is applied on the training set dataset is divided into training and testing parts. After data pre-processing, ARM is applied on the training set.

Association rules are generated based on almost 70% of included participants (i.e. the training dataset) and then clustered based on commonality of cases satisfying the rule antecedents. The remaining participants are taken as test cases to be assigned to the rule clusters to provide sets of at-risk patients sharing common MI risk factors.

3.2. Data Pre-processing

The Framingham dataset provides well-known risk factor variables such as age, sex, systolic blood pressure, total cholesterol/HDL ratio and smoking status (Mannana et al. 2012, Rodondi et al. 2012). At this stage, we focus on associations between the set of risk factors and MI (specifically, hospitalisation for an MI) with first occurrence between five to ten years after baseline data collection for Framingham participants without prior history of cardiovascular
disease. Both the training and testing sets have approximately 3% of cases with this outcome.

### 3.2.1. Feature Extraction

To test our approach we selected 10 Framingham variables; examples of participants records with extracted features are displayed in Table 1, in which, RANDID is the patient identifier, SEX="1" represents male. Span of AGE is from 32 to 81. SYSBP has range from 83.5 to 295. Range of DIABP is 30-150. CIGPDAY has maximum of 90 cigarettes per day. BMI spans from 14.43 to 56.8. Range of HEARTRTE is 37-220. And GLUCOSE has range from 39 to 478. The Framingham dataset divided participants’ attained education into four levels: “1”= 0-11 years; “2”= high school diploma; “3”= some college and “4”= college degree or more. If HOSPMI is 1, it means the participants had a first MI occurring within the period of interest (5-10 years after baseline).

### 3.2.2. Features Pre-processing

Continuous variables must be processed into discrete ranges (a process known as ‘discretization’) to form the antecedent predicates for candidate rules. However, many medical results are continuous as shown in Table 1 (e.g. BMI and SYSBP). In this experiment, in order to generate CARs that are as intuitive as possible for the clinical community, data discretization is based on the New Zealand Heart Foundation Guideline (New Zealand Guidelines Group, 2012) with resulting data ranges as shown in Table 2. Missing values is a common issue in medical data. To address this issue, we apply a simple imputation method that replaces the missing values with global means.

### 3.3. Cluster-ARM Integrated Framework

Based on the generic knowledge discovery framework mentioned above, we apply an existing ARM tool to generate and filter CARs for chronic disease outcome. CARM is applied to reveal associations for the outcome of interest (i.e., “MI within 5-10 years: yes/no” as the right-hand side, RHS) with the risk factors in the left-hand Side (LHS). Clustering techniques with different settings are then used to characterize these CARs generated on the training set, and we further evaluate these CAR Clusters on the testing set. The proposed knowledge discovery framework is shown in Figure 1.

**Figure 1: Cluster-ARM Integrated Framework**

In our proposed framework, we plan to cluster CARs by similarity of patient cases satisfying the LHS (antecedent) of each rule. Every CAR corresponds to a binary patient satisfaction vector (1 if the participant satisfies that rule’s LHS). The more patients in common between two CARs, the more likely they are to be clustered into the same group. As an illustration, in Figure 2 there are nine patients (p1 to p9) and three CARs. Each CAR has a patient satisfaction vector shown in the table at the bottom of the figure. $U$ indicates the universe of the patient dataset upon which rules are generated. By calculating the number of patients shared between two CARs, it can be seen that CAR1 and CAR2 are closer to each as compared to CAR3. By applying clustering techniques on the patient satisfaction vector data, CAR1 and 2 are most likely to be clustered together.

**Figure 2: Illustration of Clustering CARs via Patient Satisfaction Vector**

### Table 1: Examples of Participant Records with Extracted Features

<table>
<thead>
<tr>
<th>RANDID</th>
<th>SEX</th>
<th>TOTCHOL</th>
<th>AGE</th>
<th>SYSBP</th>
<th>DIABP</th>
<th>CIGPDAY</th>
<th>BMI</th>
<th>HEARTRTE</th>
<th>GLUCOSE</th>
<th>EDUC</th>
<th>HOSPMI</th>
</tr>
</thead>
<tbody>
<tr>
<td>2448</td>
<td>1</td>
<td>195</td>
<td>39</td>
<td>106</td>
<td>70</td>
<td>0</td>
<td>26.97</td>
<td>80</td>
<td>77</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>2448</td>
<td>1</td>
<td>209</td>
<td>52</td>
<td>121</td>
<td>66</td>
<td>0</td>
<td>-</td>
<td>69</td>
<td>92</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>6238</td>
<td>0</td>
<td>250</td>
<td>46</td>
<td>121</td>
<td>81</td>
<td>0</td>
<td>28.73</td>
<td>95</td>
<td>76</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>6238</td>
<td>0</td>
<td>260</td>
<td>52</td>
<td>105</td>
<td>69.5</td>
<td>0</td>
<td>29.43</td>
<td>80</td>
<td>86</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>6238</td>
<td>0</td>
<td>237</td>
<td>58</td>
<td>108</td>
<td>66</td>
<td>0</td>
<td>28.5</td>
<td>80</td>
<td>71</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Features</td>
<td>Range</td>
<td>Description</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------------------------</td>
<td>-------</td>
<td>---------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AGE</td>
<td>45-54</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>55-64</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>65-74</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>75-84</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CIGPDAY</td>
<td>0</td>
<td>Non-Smoker</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Cigarettes Smoked Per Day)</td>
<td>1-10</td>
<td>1-10 Cigarettes per Day</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>11-20</td>
<td>11-20 Cigarettes per Day</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>≥ 21</td>
<td>More than One Pack per Day</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BMI (Body Mass Index)</td>
<td>(0,25)</td>
<td>Normal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[25, 30)</td>
<td>Overweight</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>≥ 30</td>
<td>Obese</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GLUCOSE</td>
<td>&lt; 120</td>
<td>Normal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Body Mass Index)</td>
<td>[120, 220]</td>
<td>Diabetes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HEARTRATE (Heart Rate)</td>
<td>(0, 60)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[60, 100)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>&gt;100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SYSBP (Systolic Blood Pressure)</td>
<td>[0, 130)</td>
<td>Normal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>≥130</td>
<td>High</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DIABP (Diastolic Blood Pressure)</td>
<td>[0, 85)</td>
<td>Normal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>≥85</td>
<td>High</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOTCHOL (Total Cholesterol)</td>
<td>(200,239)</td>
<td>Moderate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>&gt;239</td>
<td>High</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Feature Ranges and Description

<table>
<thead>
<tr>
<th>NO.</th>
<th>RULES</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>IF SEX=MEN THEN HOSPMI=OCCURED</td>
</tr>
<tr>
<td>2.</td>
<td>IF SYSBP=HIGH THEN HOSPMI=OCCURED</td>
</tr>
<tr>
<td>3.</td>
<td>IF TOTCHOL=HIGH THEN HOSPMI=OCCURED</td>
</tr>
<tr>
<td>4.</td>
<td>IF SEX=MEN &amp; TOTCHOL=HIGH &amp; SYSBP=HIGH &amp; BMI=OVER_WEIGHT &amp; EDUC=SOME_COLLEGE THEN HOSPMI=OCCURED</td>
</tr>
<tr>
<td>5.</td>
<td>IF SYSBP=HIGH &amp; CIGPDAY=NONSMOKER THEN HOSPMI=OCCURED</td>
</tr>
<tr>
<td>6.</td>
<td>IF CIGPDAY=11_20 THEN HOSPMI=OCCURED</td>
</tr>
<tr>
<td>7.</td>
<td>IF DIABP=HIGH &amp; CIGPDAY=NONSMOKER THEN HOSPMI=OCCURED</td>
</tr>
<tr>
<td>8.</td>
<td>IF AGE=0-44 &amp; AGE=45-54 &amp; BMI=OVER_WEIGHT THEN HOSPMI=OCCURED</td>
</tr>
<tr>
<td>9.</td>
<td>IF DIABP=HIGH &amp; EDUC=0_11_YEARS &amp; EDUC=SOME_SCHOOL_DIPLOMA THEN HOSPMI=OCCURED</td>
</tr>
<tr>
<td>10.</td>
<td>IF CIGPDAY=MORE_THAN_20 THEN HOSPMI=OCCURED</td>
</tr>
<tr>
<td>11.</td>
<td>IF AGE=55-64 &amp; CIGPDAY=NONSMOKER &amp; BMI=NORMAL THEN HOSPMI=OCCURED</td>
</tr>
<tr>
<td>12.</td>
<td>IF SYSBP=HIGH &amp; EDUC=COLLEGE_DEGREE_OR_MORE THEN HOSPMI=OCCURED</td>
</tr>
<tr>
<td>13.</td>
<td>IF SEX=MEN &amp; TOTCHOL=HIGH &amp; SYSBP=HIGH &amp; EDUC=0_11_YEARS &amp; EDUC=SOME_SCHOOL_DIPLOMA THEN HOSPMI=OCCURED</td>
</tr>
<tr>
<td>14.</td>
<td>IF DIABP=HIGH &amp; EDUC=COLLEGE_DEGREE_OR_MORE THEN HOSPMI=OCCURED</td>
</tr>
<tr>
<td>15.</td>
<td>IF SEX=MEN &amp; HEARTRATE=60-100 THEN HOSPMI=OCCURED</td>
</tr>
<tr>
<td>16.</td>
<td>IF TOTCHOL=HIGH &amp; HEARTRATE=60-100 THEN HOSPMI=OCCURED</td>
</tr>
<tr>
<td>17.</td>
<td>IF GLUCOSE=DIABETES THEN HOSPMI=OCCURED</td>
</tr>
<tr>
<td>18.</td>
<td>IF AGE=65-74 THEN HOSPMI=OCCURED</td>
</tr>
<tr>
<td>19.</td>
<td>IF SEX=MEN &amp; TOTCHOL=HIGH &amp; HEARTRATE=60-100 THEN HOSPMI=OCCURED</td>
</tr>
<tr>
<td>20.</td>
<td>IF SEX=MEN &amp; GLUCOSE=DIABETES THEN HOSPMI=OCCURED</td>
</tr>
<tr>
<td>21.</td>
<td>IF CIGPDAY=MORE_THAN_20 &amp; BMI=OBESITY THEN HOSPMI=OCCURED</td>
</tr>
<tr>
<td>22.</td>
<td>IF CIGPDAY=MORE_THAN_20 &amp; BMI=OVER_WEIGHT &amp; EDUC=OVER_WEIGHT THEN HOSPMI=OCCURED</td>
</tr>
<tr>
<td>23.</td>
<td>IF SEX=MEN &amp; GLUCOSE=DIABETES &amp; BMI=OVER_WEIGHT THEN HOSPMI=OCCURED</td>
</tr>
<tr>
<td>24.</td>
<td>IF HEARTRATE=101-220 THEN HOSPMI=OCCURED</td>
</tr>
<tr>
<td>25.</td>
<td>IF AGE=55-64 &amp; CIGPDAY=MORE_THAN_20 THEN HOSPMI=OCCURED</td>
</tr>
<tr>
<td>26.</td>
<td>IF SEX=MEN &amp; TOTCHOL=HIGH &amp; Bmi=OVER_WEIGHT &amp; EDUC=SOME_COLLEGE THEN HOSPMI=OCCURED</td>
</tr>
<tr>
<td>27.</td>
<td>IF SEX=MEN &amp; TOTCHOL=MODERATE &amp; SYSBP=HIGH &amp; HEARTRATE=60-100 &amp; EDUC=0_11_YEARS &amp; EDUC=SOME_SCHOOL_DIPLOMA THEN HOSPMI=OCCURED</td>
</tr>
</tbody>
</table>

Table 3: Selected CARs

This section presents the implementation of the proposed framework on the Framingham dataset with results. Data preparation is explained in section 3 where data is divided into training and testing sets. We use the training set to generate and cluster CARs and then evaluate CAR clusters on the testing set by analysing the ‘significance’ (degree of MI risk and number of cases covered) and independence (non-overlap) of patient segmentations by CAR clusters.

4.1. CARs Mining and Filtering

We use Brute (Riddle et al., 1994) for CARM generation on the training dataset. Brute is a rule exploration tool using an exhaustive and depth-bounded search of the space of decision rules with rule filtering on chi-square values [12]. We decided to explore rules with a maximum of seven conjunctions on the LHS. 170 CARs were generated. Lift and Coverage are used for selecting interesting rules. 27 CARs, shown in Table 3, were selected with Lift greater than 3.2 as well as Coverage greater than 7.

4.2. CAR Clustering

For the current experiment, SAS software (SAS Institute Inc, 2013) is used for CAR clustering based on dissimilarity of participant coverage on the training dataset. We used the Jaccard distance to define degree of individuality in participant covered for any two CARs as normalised by the total number of participants covered by the two CARs. Its definition is shown as follows:

$$J_\delta(A,B) = \frac{|A \cup B| - |A \cap B|}{|A \cup B|}$$

Where $A$ and $B$ are the patient satisfaction vectors of the two CARs. The higher the Jaccard distance between two CARs, the more dissimilar two CARs are. We have applied four hierarchical clustering methods: Average, Complete, Single and Ward’s. We specified for the methods to return 3, 4 or 5 clusters — figuring these to be manageable numbers of patient sets for future review by healthcare experts, and representing a considerable reduction from the total 27 CARs. CAR clustering results are displayed in Table 4, in which CARs are represented by CAR index number as per Table 3.

4.3. Patient Segmentation

We then evaluate CAR clusters by assessing significance (a combination of cluster size in terms of patients mapped to the cluster and degree of relative risk of MI for those patients) and independence of CAR cluster-based patient segmentation (i.e. striving for minimal overlap of patients between CAR clusters) on the testing set. Based on the previous steps,

<table>
<thead>
<tr>
<th>3 CAR Clusters</th>
<th>4 CAR Clusters</th>
<th>5 CAR Clusters</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2,3,5,6,7,8,9,10,11,12,13,14,15,16,19,24,26,27</td>
<td>1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,19,21,22,24,25,26,27</td>
</tr>
<tr>
<td>2</td>
<td>17,18,20,23,10,21,22,25</td>
<td>17,20,23</td>
</tr>
<tr>
<td>3</td>
<td>4,15,16,19,24,26,27,17,18,20,23</td>
<td>18</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>4 CAR Clusters</th>
<th>4 CAR Clusters</th>
<th>5 CAR Clusters</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2,3,5,6,7,8,9,10,11,12,13,14</td>
<td>1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,22,25,26,27</td>
</tr>
<tr>
<td>2</td>
<td>17,20,23</td>
<td>17,20,23</td>
</tr>
<tr>
<td>3</td>
<td>4,15,16,19,24,26,27</td>
<td>17,18,20,23</td>
</tr>
<tr>
<td>4</td>
<td>18</td>
<td>4,15,16,19,24,26,27,15,16,19,24</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>5 CAR Clusters</th>
<th>5 CAR Clusters</th>
<th>5 CAR Clusters</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2,3,5,6,7,8,9,10,11,12,13,14</td>
<td>1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,26,27</td>
</tr>
<tr>
<td>2</td>
<td>17,20,23,10,21,22,25</td>
<td>17,20,23</td>
</tr>
<tr>
<td>3</td>
<td>15,16,19,24</td>
<td>17,20,23</td>
</tr>
<tr>
<td>4</td>
<td>4,26,27</td>
<td>4,15,16,19,24,26,27</td>
</tr>
<tr>
<td>5</td>
<td>18</td>
<td>18</td>
</tr>
</tbody>
</table>

Table 4: CAR Clusters with Different Settings
the 27 CARs are clustered by 12 clustering settings with results as per Table 4. These 12 sets of CAR clusters are then each applied to segment the testing set of participants.

The number of test cases for segmentation is decided by the number of CAR cluster (one set of patients per CAR cluster). There remains the question, however, of how to map test cases to CAR clusters, or in other words, to determine if a CAR cluster ‘covers’ a test case. Any given test case may satisfy the LHS of zero, one or more of the CARs in a given cluster. For each CAR cluster we allow a different threshold of CAR satisfaction for mapping test cases to that cluster. Test cases are mapped to a CAR cluster when they satisfy the LHS of that cluster’s ‘satisfaction level’ number of CARs. Precision and recall are used to guide the search for each CAR cluster’s satisfaction level. In our experiments, precision represents the fraction of covered test cases that had an MI (the frequency with which the participants mapped to the cluster have the adverse event); recall represents the fraction of test cases with MI in the test set that are covered by (mapped to) the CAR cluster. These two performance measures are combined into an F-measure, which is defined conventionally as:

$$F_{\beta} = \frac{(1 + \beta^2) \times p \times r}{\beta^2 \times p + r}$$

Where $\beta$ controls the relative weight of precision and recall ($\beta=1$ is used for equal contribution); $p$ stands for precision and $r$ is recall. We believe that the significance of CAR clusters is dominantly reflected by high recall and assign $\beta=2$.

For each cluster, the threshold CAR satisfaction level is determined by the highest F-measure. This is demonstrated in Table 5 for the first cluster derived by the Complete method for the 5-cluster solution. The first column of Table 5 indicates the minimum number of CARs from this cluster that test cases have to satisfy in order to be mapped to the cluster. In the second column, ‘# of participants covered’ shows the number of test cases mapped to this cluster at that satisfaction level. For example, there are 979 test cases that satisfy at least one CAR in this CAR cluster, and this includes 34 participants with an MI event. However, there are only 15 test cases (with one participant with an MI event included) assigned to this cluster if we raise the CAR satisfaction level to 8.

Each CAR satisfaction level has one F-measure value and the satisfaction level with highest F-measure is decided as the threshold to be applied. According to highest F-measure (i.e. 0.24), test cases can be mapped to this CAR cluster if they satisfy at least four CARs in this cluster. Each participant is allowed to be mapped to more than one cluster; i.e. while the CARs themselves are partitioned into clusters, the test case sets associated with those CAR clusters may (and generally will) have a degree of overlap in membership.

Table 6 provides a summary of the best F-measures and relative risk (cluster precision for MI divided by overall rate of MI in the test data) for the test cases associated with the CAR clusters of each of the 12 experimental runs (four clustering methods × three-, four- or five-cluster solutions). The first column identifies the cluster; for instance, “4-3” denotes the third cluster in the 4-cluster solution. #Cases indicates the number of test cases assigned to the cluster and #MI indicates the number of participants with an MI event included in #Cases.

<table>
<thead>
<tr>
<th>Satisfaction Level</th>
<th># of Participants Covered</th>
<th>Covered Participants with MI</th>
<th>Precision</th>
<th>Recall</th>
<th>F-measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>979</td>
<td>34</td>
<td>0.035</td>
<td>1</td>
<td>0.15</td>
</tr>
<tr>
<td>2</td>
<td>726</td>
<td>32</td>
<td>0.04</td>
<td>0.94</td>
<td>0.18</td>
</tr>
<tr>
<td>3</td>
<td>429</td>
<td>25</td>
<td>0.06</td>
<td>0.73</td>
<td>0.22</td>
</tr>
<tr>
<td>4</td>
<td>259</td>
<td>19</td>
<td>0.07</td>
<td>0.566</td>
<td>0.24</td>
</tr>
<tr>
<td>5</td>
<td>157</td>
<td>12</td>
<td>0.08</td>
<td>0.35</td>
<td>0.20</td>
</tr>
<tr>
<td>6</td>
<td>74</td>
<td>7</td>
<td>0.09</td>
<td>0.2</td>
<td>0.17</td>
</tr>
<tr>
<td>7</td>
<td>36</td>
<td>5</td>
<td>0.14</td>
<td>0.15</td>
<td>0.15</td>
</tr>
<tr>
<td>8</td>
<td>15</td>
<td>1</td>
<td>0.07</td>
<td>0.03</td>
<td>0.03</td>
</tr>
<tr>
<td>9</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 5: Example of Defining Cluster Satisfaction Level for a CAR Cluster with 11 CARs
<table>
<thead>
<tr>
<th>CAR Cluster</th>
<th>AVERAGE</th>
<th>COMPLETE</th>
<th>SINGLE</th>
<th>WARD's</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Relative Risk</td>
<td>F-measure</td>
<td>#Cases</td>
<td>#MI</td>
</tr>
<tr>
<td>3-1</td>
<td>2.3</td>
<td>0.19</td>
<td>276</td>
<td>20</td>
</tr>
<tr>
<td>3-2</td>
<td>2.0</td>
<td>0.08</td>
<td>63</td>
<td>4</td>
</tr>
<tr>
<td>3-3</td>
<td>1.7</td>
<td>0.08</td>
<td>76</td>
<td>4</td>
</tr>
<tr>
<td>4-1</td>
<td>2.3</td>
<td>0.19</td>
<td>276</td>
<td>20</td>
</tr>
<tr>
<td>4-2</td>
<td>2.4</td>
<td>0.05</td>
<td>27</td>
<td>2</td>
</tr>
<tr>
<td>4-3</td>
<td>1.7</td>
<td>0.08</td>
<td>76</td>
<td>4</td>
</tr>
<tr>
<td>5-1</td>
<td>2.3</td>
<td>0.24</td>
<td>276</td>
<td>20</td>
</tr>
<tr>
<td>5-2</td>
<td>2.4</td>
<td>0.06</td>
<td>27</td>
<td>2</td>
</tr>
<tr>
<td>5-3</td>
<td>2.0</td>
<td>0.10</td>
<td>64</td>
<td>4</td>
</tr>
<tr>
<td>5-4</td>
<td>5.4</td>
<td>0.04</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>5-5</td>
<td>1.7</td>
<td>0.06</td>
<td>37</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 6: Comparison of CAR Clusters for Participant Segmentation
All clustering settings yield a large cluster covering at least 200 test cases with the largest or second-largest number of CARs. They all share the same CAR satisfaction level: if the test case satisfies four or more CARs of this largest CAR cluster, the test cases will be associated with that cluster. On the other hand, the threshold CAR satisfaction level for all the rest of the clusters is one, except for the 5-cluster solution with the Average method, for which the satisfaction level is two CARs.

In this experiment, each participant in the testing set is allowed to have more than one CAR cluster assignment; i.e. for a certain clustering setting, one test case may belong to two or more clusters at the same time. While some such overlap is acceptable, we believe that excessive overlap could reduce the intuitiveness and value of the test case sets for expert interpretation and have chosen a criterion to evaluate the overall goodness of a CAR cluster solution using the Ochiai coefficient (Ochiai, 1957) in the denominator. The Ochiai coefficient can be regarded as a cosine similarity when the data type is a bit vector; its definition is:

$$K = \frac{n(A \cap B)}{\sqrt{n(A) \times n(B)}}$$

Where \(n(A)\) and \(n(B)\) indicate the number of covered test cases in any two clusters of a certain clustering setting. And \(n(A \cap B)\) indicates the number of overlapping test cases between the two clusters.

The goodness of individual CAR clusters relies on high risk participants being covered, which was achieved by the F-measure. Hence, we evaluate a set of CAR clusters (the ‘solution’ of 3, 4 or 5 clusters from a single experimental run) by both F-measure and Ochiai coefficient. The evaluation definition is:

$$V_{ij} = \frac{F_i + F_j}{K_{ij}} \quad (i \neq j)$$

$$\bar{V} = 2 \sum_{i \neq j} V_{ij} \frac{n_i n_j}{n(n-1)}$$

where \(F_i\) and \(F_j\) are the F-measures of clusters \(i\) and \(j\) in a given clustering setting and \(V_{ij}\) is the fractional contribution of the pair \(i,j\) to the solution’s value; \(n\) indicates the number of clusters (\(n \in \{3, 4, 5\}\)). \(\bar{V}\) represents the mean of \(V_{ij}\) for a CAR cluster set based on all pairs of clusters and is taken as the overall value of the solution. Comparison of \(\bar{V}\) in different clustering settings is shown in Table 7; the highest \(\bar{V}\) for each method is underlined. The highest \(\bar{V}\) for the Complete method, with five CAR clusters, is greater than all rest: it is taken as the best clustering setting in the current experiment. We then analyze the mean and standard deviation of test cases assigned to the CAR clusters of this best solution with results shown in Table 8.

<table>
<thead>
<tr>
<th>No. of Cluster</th>
<th>Average</th>
<th>Complete</th>
<th>Single</th>
<th>Ward’s</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>2.62</td>
<td>2.62</td>
<td>2.58</td>
<td>1.48</td>
</tr>
<tr>
<td>4</td>
<td>2.04</td>
<td>2.7</td>
<td>2.68</td>
<td>2.</td>
</tr>
<tr>
<td>5</td>
<td>2.1</td>
<td>2.8</td>
<td>2.4</td>
<td>2.26</td>
</tr>
</tbody>
</table>

Table 7: Comparison of \(\bar{V}\) on CAR Clusters for Patient Segmentation

5. Discussion

Overall, the high relative risk of all clusters in Table 6 indicates that our CAR-clustering framework is successful in revealing groups of participants at high risk (at least elevated as compared to the average risk for a case in the dataset) for an adverse event. The technique is helpful for managing the proliferation of CARs inherent in ARM in that we were able to reduce the set of discovered CARs (still 27 after interestingness reduction based on Lift and Coverage) to a user-specified number of CAR clusters and associated largely-non-overlapping patient sets. While some clustering settings yielded very high risk groups (e.g. 400%+ for the 4th cluster for the 5-cluster solution with the Average method, and the 4th cluster of 5 with Ward’s method), these clusters provide relatively low participant coverage. Our evaluation function has balanced precision, recall and overlap to select a clustering solution that provides several substantial participants groups, each with high relative risk and each due to different complexes of risk factors.

Looking at the 5-cluster solution with the Complete method in Table 8, we note across all clusters a high prevalence of history of hypertension and a mean BMI in the ‘overweight’ range, and with the first four clusters majority male. By referring to the CAR details from Table 3 and CARs-by-cluster in Table 4, we can get a sense of the dominant characteristics of some of the clusters. For instance: Cluster 5, based on just CAR 18, is older participants; Cluster 3 (based on CARS 17, 20 and 23) has diabetes as a dominant factor and shows greatly elevated mean glucose as compared to the other clusters; and Cluster 2 (based on CARS 10, 21, 22 and 25) is made up of heavy smokers. As such, each aligns neatly to a powerful cardiovascular disease risk factor (age, diabetes and smoking, respectively). Clusters 1 and 4 involve a larger number of CARs and thus, unsurprisingly, are more complex to interpret.
Table 8: Characteristics of Participant Sets for 5-Cluster Solution

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Cluster</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean / freq</td>
<td>Mean / freq</td>
<td>Mean / freq</td>
<td>Mean / freq</td>
<td>Mean / freq</td>
<td>Mean / freq</td>
</tr>
<tr>
<td>N</td>
<td>259</td>
<td>130</td>
<td>27</td>
<td>76</td>
<td>37</td>
<td>37</td>
</tr>
<tr>
<td>HOSPMI</td>
<td>0.07</td>
<td>-</td>
<td>0.06</td>
<td>-</td>
<td>0.07</td>
<td>-</td>
</tr>
<tr>
<td>MALE</td>
<td>0.66</td>
<td>-</td>
<td>0.76</td>
<td>-</td>
<td>0.55</td>
<td>-</td>
</tr>
<tr>
<td>TOTCHOL</td>
<td>253.80</td>
<td>3.18</td>
<td>233.20</td>
<td>3.70</td>
<td>231.60</td>
<td>7.27</td>
</tr>
<tr>
<td>AGE</td>
<td>51.05</td>
<td>0.48</td>
<td>46.15</td>
<td>0.62</td>
<td>51.07</td>
<td>1.70</td>
</tr>
<tr>
<td>SYSBP</td>
<td>146.10</td>
<td>1.08</td>
<td>128.70</td>
<td>1.57</td>
<td>141.00</td>
<td>4.66</td>
</tr>
<tr>
<td>DIABP</td>
<td>90.46</td>
<td>0.62</td>
<td>81.78</td>
<td>0.91</td>
<td>85.01</td>
<td>2.56</td>
</tr>
<tr>
<td>HEARTRTE</td>
<td>78.67</td>
<td>0.74</td>
<td>78.53</td>
<td>0.98</td>
<td>82.81</td>
<td>2.44</td>
</tr>
<tr>
<td>GLUCOSE</td>
<td>80.16</td>
<td>1.07</td>
<td>78.30</td>
<td>1.41</td>
<td>181.70</td>
<td>13.88</td>
</tr>
<tr>
<td>CIGPDAY</td>
<td>15.86</td>
<td>0.77</td>
<td>34.00</td>
<td>0.65</td>
<td>26.80</td>
<td>0.92</td>
</tr>
<tr>
<td>BMI</td>
<td>26.87</td>
<td>0.23</td>
<td>25.61</td>
<td>0.33</td>
<td>26.00</td>
<td>0.92</td>
</tr>
<tr>
<td>HYPERTEN</td>
<td>0.90</td>
<td>-</td>
<td>0.69</td>
<td>-</td>
<td>0.81</td>
<td>-</td>
</tr>
<tr>
<td>Relative Risk</td>
<td>2.4</td>
<td>2.0</td>
<td>2.4</td>
<td>1.7</td>
<td>1.7</td>
<td>1.7</td>
</tr>
</tbody>
</table>

This study has a number of inherent limitations. Notably, the Framingham ‘teaching’ data set utilised herein, due to the nature of its anonymisation methods, is not deemed suitable for deriving population health statistics for real populations. We make no assertion that the specific clusters as per Table 8 will emerge in similar form based on other population data. While the Complete method (complete linkage, ‘furthest neighbour’ clustering; SAS Institute, 2013) gave the best performance in our case, this may not continue to dominate in performance for other chronic disease related adverse event data. Moreover, while we have illustrated a method to manage the proliferation of ARM output for adverse event prediction, we have not demonstrated that it is a superior method to other candidates. For instance, Ordonez et al. (2006) have pursued similar objectives with respect to rule proliferation management in the area of ARM for heart disease using a process to ‘summarize’ rules that have the same consequents. Alternatively, a more direct approach would have been simply to cluster the patients on their attributes without an intermediate CAR derivation process, similar to what is commonly done in clustering for consumer relationship management (CRM; Collica, 2007); such an approach, however, would not provide descriptive rules to characterise each cluster. Areas for future research are outlined in the final section below.

6. Conclusions and Future Research

In this paper, we have explored the problem of managing the proliferation of association rules from ARM to provide insight in population patterns of potentially preventable chronic disease related adverse events. A Clustered-CARs based framework was proposed to manage the large number of discovered rules and help identify and characterise groups of at-risk patients. Using the Framing Heart Study teaching data set we demonstrated the ability to identify patient sub-groups with low overlap, substantial membership and elevated relative risk as compared to the overall dataset. We explored a range of clustering settings finding the Complete method (complete linkage, ‘furthest neighbour’) to obtain the CAR clusters with the best balance of precision, recall and low overlap on these specific data. Our framework was successful in substantially managing down a large set of association rules (27 after conventional interestingness filtering) to a manageable set of clusters providing distinctive at-risk patient sub-group characterisations.

Further experiments will analyse the full Framingham Heart Study cohort and New Zealand VIEW data (under human research ethics protocols that have been obtained). A particular challenge and opportunity in these data is the greater range of variables and temporal patterns, particularly around health service utilisation patterns (including laboratory tests and medication dispensing) allowing us to generate features based on physiological changes and treatment patterns. The domain experts from the VIEW programme (epidemiologists, cardiologists, etc.) will be able to provide feedback on the face validity and preventative care potential of the discovered patterns.

Acknowledgments

We would like to thank our colleagues in the VIEW programme, particularly Prof Rod Jackson and A/Prof Roger Marshall, as well as Prof Pat Langley in Computer Science, for their helpful feedback on our approach.

7. References

AIHW (2009): Australian Hospital Statistics 2007-08, Canberra, AIHW.


Biologic Specimen and Data Repository Information Coordinating Center, https://biolincc.nhlbi.nih.gov/static/studies/teaching/framdoc.pdf


SAS Institute Inc (2013), the data analysis for this paper was generated using SAS software, Version 9.3 of the SAS System for Win7. Copyright 2013. SAS Institute Inc. SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc., Cary, NC, USA.


Building an Empirical Treatment Protocol from High-Resolution Traumatic Brain Injury Data

ANTHONY STELL¹, LAURA MOSS¹,²,³, IAN PIPER¹,³
¹Dept of Clinical Physics, University of Glasgow, Glasgow, UK
²Dept of Computing Science, University of Aberdeen, Aberdeen, UK
³Dept of Clinical Engineering & Bioengineering, NHS Greater Glasgow & Clyde, Glasgow, UK
a.stell.1@research.gla.ac.uk

Abstract
An informatics issue common for many fields of medical research is the poor standardisation of baseline clinical management data, which can have a large negative impact on the statistical power of drug studies making use of that data. This baseline variation can be for many reasons – e.g. rarity of the condition – but, despite the development of standardised medical guidelines in many areas, it is still often the case that study data is affected by the “real world” differences in treatment protocols. To improve understanding of that management baseline in general, this paper describes work that builds up an empirical treatment pattern from retrospective intensive care unit (ICU) data. The ultimate goal is to build protocol “objects” that can be compared between specialist centres or “gold standard” guidelines. Variation and differences between these objects can then be quantified - and potentially mitigated - to allow a more standardised comparison of data for studies, as well as providing information on audit and guideline adherence. From a combination of event detection from high-resolution physiological output and association of those detected events with annotated treatment information, an empirical data-driven notion of treatment protocols across specialist centres can be built. Using data drawn from Traumatic Brain Injury (TBI) studies, the initial steps of this technological work – including the algorithms and assumptions of these two key functions – are presented. The results when applied to a specific TBI data-set (Piper et al 2010) show how the event numbers vary when key parameters are changed (e.g. the hold-down time) and how this impacts clinical decisions and trial conduct.

Keywords: treatment-protocol, study data, event-detection

1 Introduction
Certain areas of medical research suffer from low statistical power – the ability to identify a statistically significant result – in drug and intervention trials and studies. There are many reasons that can contribute to this: the rarity of the medical conditions; the complexity of the organ affected; etc. Often the result of this low power is that trials must recruit from larger population distributions and necessarily involve more centres that specialise in the condition in question.

When this occurs, a large source of variation now includes the differences in treatment protocol between those specialist centres. A move to standardise the administration of treatment procedures has gained ground over recent decades, resulting in the widespread development and adoption of clinical guidelines (Woolf et al 1999). Following a clinical guideline allows clinicians to follow reproducible treatment procedures in a standard manner. Whilst providing the best available information on reproducible care, this standardisation in treatment and care also helps progress medical research using the bedrock of the scientific method: refine and improve treatment by understanding the current environment then vary one parameter at a time and monitor the effects.

Despite this significant advance in medical treatment provision, it is still the case that guidelines are not always followed. This disparity can occur for many reasons but a common and important issue that has a subtle impact on the inputs to trial and study data, is the difference between reported and actual treatment or care actions.

It is this difference that the overall goal of this work will attempt to capture technologically, by analysing retrospective study data in the Traumatic Brain Injury domain (TBI). TBI is a prime example of a medical research area with an abundance of low-power trial data. It is widely acknowledged that the progress in understanding putative drug treatments for TBI has been greatly hindered by this lack of useful trial data (Lu et al 2012). The data provides little insight into effective treatment data because of the complex medical processes involved in TBI, which means that any trial must recruit large patient numbers, and therefore has to recruit from widely distributed areas. The result is an overall lack of confidence in treatments and interventions for TBI – most explicitly noted in the lack of certainty underpinning the recommendations made by the authoritative guidelines in the space, made by the Brain Trauma Foundation (BTF) (Bullock et al 1996). It should also be noted that these issues of poor trial data are certainly not exclusive to TBI and exist in other areas such as adrenal cancer where the rarity of the condition means that trials must recruit globally and similarly try to use novel analysis techniques to gain useful insight (Schteingart 2005).

To pursue a solution to this general issue, a worthwhile endeavour is to analyse the data-sets of clinical trials and other studies that already exist and attempt to build patterns of treatment protocols that can then be re-used for future trials. The technological proposal outlined here is an approach for the compilation
and subsequent comparison of clinical workflows between specialist centres, clinicians and patients. This approach can be broken down into the following stages:

1. Detection of events from physiological ICU time-series data
2. Association of those events with treatment information
3. Compiling these associations into a pattern of treatment (a “protocol object”) that can be expressed in a standard manner
4. Comparison of these protocol objects between centres within study data-sets (and outside those data-sets for validation)
5. Quantifying the differences that are detected
6. Mitigating or accounting for these differences so that the inputs to trial or studies can be more fairly understood. This is the critical aim of the work – mitigating these differences will potentially allow trial data to show more statistically significant findings.

This paper describes the development and implementation of an algorithm to detect events and associate corresponding treatment from patient data (steps 1 and 2 in the list above).

2 Background

2.1 Trial data

Poor trial data is an issue that potentially affects all areas of medical research. In the specific domain of TBI, there is a general acknowledgement that trial data lacks the statistical significance to move the understanding of treatments forward (Lu et al 2012). Several Cochrane reviews (systematic reviews that analyse a collection of studies to draw additional insight) have been conducted and their findings are inconclusive (often contradicting the original “assumed” clinical finding, such as the use of barbiturates therapy in TBI (Roberts and Sydenham 1999)). It is the case that meta-analyses are only as effective as the studies that they collectively review. Therefore, if those studies suffer from bad design or poor numbers and detail, then a meta-analysis won’t highlight anything new. The IMPACT project is an example initiative that attempts to solve the resulting problem using statistical techniques (Maas et al 2010), by modifying the outcome information into more detailed categories and specifically surveying the strictness of enrolment criteria. The results of this are that statistical efficiency is improved by 40%. However it is acknowledged that further validation of these results and an investigation of alternative methods is required. Differences in the baseline clinical management have been quoted as a primary concern in the lack of significant study output (Lingsma et al 2011), so attempts to analyse the nature of standardisation in this area is worth pursuing.

To investigate differences in baseline data it is instructive to first understand the work conducted so far in standardising general treatment and other study protocols. This is best done by looking at the authoritative clinical guidelines in the domain. In TBI, these are the guidelines compiled and maintained by the Brain Trauma Foundation (BTF), which cover all types of situations including intensive care stays, emergency accident-scene care and other specific situations such as trauma sustained whilst in military combat (Bullock et al 1996). In recognition of the varying certainty of the evidence behind their effectiveness, the BTF guidelines provide a tabulation of the confidence level behind a specific recommendation. Three broad classifications of guidelines are published by the BTF (in decreasing order of certainty): Standards, Guidelines and Options. The classification of a specific guideline is based on the classification of the supporting evidence: level 1, 2 and 3 treatment recommendations, supported by class 1, 2 and 3 evidence respectively (and again in decreasing order of certainty of efficacy). Surveying the TBI online searchable guidelines, it appears to be the case that there are very few level 1 treatment recommendations, and therefore a corresponding lack of standards (Shafi et al 2008). This appears to be especially the case for ICP monitoring, a particularly invasive treatment, which unfortunately has been identified as one of the primary avenues of potential progress in the treatment of TBI.

The use of these guidelines has been demonstrated to be clinically effective (Faul et al 2007), but controversy on their utility still exists (Pascual et al 2011) and much work remains in providing broad agreement in their recommendations. A recent alternative viewpoint to guideline-based treatment is a move to set up research infrastructures that support personalized medicine (Saatman et al 2008). This approach attempts to embrace the variation in data, dealing with patient information on a case-by-case basis, though it is hard to see how this can be expanded to more generalized solutions without establishing what how the individual differences between patients arise.

2.2 Technical background

The technical details of what is proposed in this work require the understanding of events in the context of physiological monitoring. Work in this area has focused on the definition of a physiological event through the Edinburgh University Secondary Insult Grade (EUSIG) (Jones et al 1994). These focus on the specific details of what physiological values should be used for threshold crossing (e.g. a value of greater than 100 beats per minute (bpm) for heart rate) and hold-down times (e.g. greater than 100 bpm for 10 minutes), but do not necessarily cover all the clinical definitions that are accepted (Donald et al 2012). Hence a valuable area to investigate is to look at the spread of these definitions and how they are represented in output ICU data. Detecting such physiological monitoring events in the context of high-resolution ICU data is a concept well-represented in the commercially available systems that can be found in modern ICU centres. Systems such as ICMS (Smielewski et al 2005), Philips CareView (Philips 2013) and Datex Ohmeda (GE Healthcare 2013) are built upon algorithms similar to the event representations referred to above, some of which have the ability to actively vary threshold warnings in response to what the favoured clinical
definition is at a particular centre (Otero et al 2009). Therefore, the data analysed using these event definitions and sourced from these and similar systems, are highly relevant to the analysis.

In terms of associating treatment annotations to detected events, causal association is a very difficult problem that requires much contextual information to unambiguously establish. High specification of the data-set used is the ideal method (e.g. a clinician directly highlighting what event they are administering the treatment for) but it is often the case that such specification is not available (Enblad et al 2004). Work has been conducted to attempt to mathematically attach a treatment to a particular event, but these necessarily have an element of probability, and hence a confidence measure (Sackarelles et al 2010). Therefore, any method that tries to establish this association can only do so to a certain degree of certainty.

In regard to the later steps outlined in section 1, there are many technological initiatives that have a potential impact on this area of evaluating data profile and clinical work-flow differences. Those that most closely link to the idea proposed (of building complex protocol objects) are clinical work-flow systems such as ProForma (Sutton and Fox 2003), which allow guidelines to be compiled and “run” (or enacted) in a programming environment. Similarly, various standards exist for describing the entities that would be required for electronic representations of clinical guidelines (Boxwala et al 2004). The work presented in this paper uses the terminology of object-oriented programming (“instantiation”, “attributes”, etc) but literature in this domain appears to have no analogy to such an approach. The ultimate goals of creating complex protocol objects also require further investigation of the best ways to parameterise, compare and measure the similarities of such objects. Initial searches in this area have highlighted the possibilities of using measures of semantic similarity (Pederson et al 2007) or using a case-based reasoning approach to establishing patient data profile similarity (Kumar et al 2009).

3 Key technical requirements

The first two functions detailed in the list in section 1 – event detection and association of treatment information – have their own set of specific requirements that are now outlined below.

3.1 Event detection

The requirements to detect events in a set of high-resolution time-series data are 1) an understanding of the structure of the event that is being detected (i.e. the object), and 2) the specific numerical knowledge that populate that structure (i.e. the instantiation of an object for a particular domain). In the case of an event, the key structural characteristics are:

• Event threshold for event start and finish – the value which acts the trigger for knowing when an event may have started
• Event hold-down – the time beyond the threshold that indicates when an event has unambiguously occurred
• Clear hold-down – the time below the threshold that indicates when an event has unambiguously finished
• Duration
• Value range

Figure 1 shows a schematic of a single physiological monitoring event, with a time-window for treatment overlaid (see section 3.2 for discussion of time windows).

![Figure 1: event definition for a given time-series physiological output. A threshold crossed for a specific period (the hold-down) indicates that an event has started. Clear hold-down indicates that the event has finished. Also shown are a treatment at a specific time-point and a time window overlaid for association of that treatment with the event.](image)

To use the object-oriented programming analogy, this event object can be thought of as a complex structure with various attributes. In the context of this work, it is very likely that the structural details of the defined object will never change, as this is a generally accepted definition of an event throughout medical literature (Donald et al 2012). This makes it a re-usable pattern ideal for use in searching physiological time-series data.

In terms of the numerical content for the TBI domain, preliminary analyses suggest that when looking at key medical parameters – intracranial pressure (ICP) and cerebral perfusion pressure (CPP) – optimum values of 20mmHg and 60mmHg respectively serve as the most popular thresholds for specialist neurosurgical centres (Jones et al 1994).

With these key pieces of structural and numerical information about event definition, an analysis program has been built that detects this event pattern within the data-set and compiles related metrics, such as event numbers, distribution, and duration. By varying the key numerical input information (e.g. move event threshold value up or down), these metrics will change and provide information about the overall clinical situation.

Though not addressed in this paper, a further significant refinement to this pattern-matching requirement is to recognise the common characteristic of extended periods of volatility in monitored physiological parameters.
output (groups of events). This is discussed further in section 6.

3.2 Association of treatments with events

The second key function is the association of treatment information with the events detected from the physiological data. A feature common to nearly all modern high-resolution ICU data-sets is the annotations of treatments administered to a patient during their stay in intensive care, such as a nurse administering analgesics to provide pain relief or a ventilator machine being attached to a patient to allow steady assistance of breathing.

As shown by those two examples, the structure of a treatment object varies depending on the nature of the treatment and can be simple or complex. The simplest form of representing a treatment would be a timestamp, and a dosage of a certain amount of drug. A more complex object would be the attachment of the ventilator, which has start and end points, duration, and a range of values depending on the breathing assistance given. Other features could also be added to these lists (thereby increasing the complexity).

For the purposes of the analysis described here, the treatment information has been reduced to the simplest point-like structure possible, consisting of only a timestamp, a value and a label. Where the treatments have more complex structures, the treatment information has been deconstructed to use the start and end points as the individual timestamps. The long-term view however is to develop the analysis to incorporate more complex definitions of treatments.

Therefore, the method of association presented involves noting a simple treatment point in the timeline associated with the physiological data. The event detection algorithm is applied to the physiological dataset and for each event detected, several time-windows differing in length (30, 60, 90 and 120 minutes) are explored around the event to identify a corresponding treatment. The first treatment found in this time window is assumed to be in response to the event that the time window is associated with. Limitations to this approach include: that there may be more than one treatment falling within the time window, relating to that event; or overlapping event time-windows may confuse the particular association of a treatment with an event. These limitations are deemed to be acceptably negligible for the analysis run so far (estimated to be less than 1% of the overall event numbers), but must be incorporated as the work progresses.

Association between two events and actions can be calculated in many ways (see section 2). In this context, the association being discussed is causal (we are attempting to establish where a treatment was applied in response to a particular event). If the data is well annotated with treatment information then an explicit parameter target (though not necessarily the exact event) will have been noted, but this level of detailed treatment information is not always available in medical data-sets. Also, the primary goal of the work is to establish treatment protocol independent of the input from the clinician themselves, so the ideal situation is where the treatment target is established without explicit direction from the clinician. However as discussed in section 2, association is an open (and largely unsolved) research question – in this context, the assumptions made to establish the association between treatment and events are as good as can be enacted in this context and time. This is an area of work that could be followed up as a separate avenue of research.

4 Method

This section describes the relevant clinical input used and the algorithms constructed to detect events and associate treatments.

4.1 Numerical instance input

As described in the previous section, for the purposes of detecting physiological monitoring events, the structural information describing such an event will remain unchanged. To implement the methods on real data-sets, the attributes listed in section 3.1 were populated with varying numerical information depending on the instantiation of the event objects. By varying the metrics in this way, information can be derived about what definitions of ICP and CPP events are most commonly used by clinicians.

Following from the most clinically relevant definitions of ICP and CPP events (see section 2), eight parameter definitions are used to cover the most likely definitions. The key point about these parameters is that the threshold value and directions indicate when an event has started or finished (noted in italicized brackets). For instance, the first definition indicates that a raised ICP event will have considered to be started once the ICP goes above 10 mmHg.

- Raised ICP #1 (> 10 mmHg)
- Raised ICP #2 (> 15 mmHg)
- Raised ICP #3 (> 20 mmHg)
- Raised ICP #4 (> 25 mmHg)
- Raised ICP #5 (> 30 mmHg)
- Lowered CPP #1 (< 50 mmHg)
- Lowered CPP #2 (< 60 mmHg)
- Lowered CPP #3 (< 70 mmHg)

The other required instance data points are the hold-down and clear hold-down times of an event. These represent the time period after a threshold-crossing where the output continues to remain above that threshold and therefore an event can be considered to have unambiguously occurred. The same method (with opposite polarity) applies at the end of an event – known as the clear hold-down time – to indicate when an event has unambiguously ended. Again, this is an attribute that remains structurally constant, but the numerical value of which can be (and is) varied according to different clinical specialists. In a similar fashion to the threshold value definitions – beginning with the most clinically relevant definitions – four values are applied representing the differences in hold-down and clear hold-down times.
These are 5 mins, 10 mins, 15 mins and 20 mins. Therefore, there are a total of 32 (8 * 4) ways that a physiological monitoring event can be detected in a data-set.

The last attribute requiring numerical variation is the time-window that provides the period over which the detected event can be associated with an annotated treatment.

The basis for the size of the time-window has been the specified reaction times for clinicians in an ICU setting (i.e. administering a drug in response to an event can be reasonably expected to be around 30 minutes). However, a large uncertainty occurs in this variable as the actual administration time can vary to a great degree from the administration reporting time – a doctor saving a patient’s life was too busy saving their life, rather than reporting and annotating the treatment). Therefore, the time-window post-event can be varied anywhere from 30 minutes to 2 hours, a value established by surveying the clinicians that contributed to the data-set (Enblad et al 2004). Extending the time-window before the event is a possibility considered due to reporting discrepancies, but the same survey established that pre-event administration reporting was unlikely. Therefore the four time-window definitions (starting at the point of event start) are: 30 mins, 60 mins, 90 mins and 120 mins. With the four time-windows, the total number of analyses for every pass of the data becomes 128 (32 * 4).

The results of this association approach are simple number counts of the treatments that fall within those time windows. Other metrics that can be compiled using this approach include measuring the time to an associated treatment. The time of all the treatments from their associated event start are listed, and the mean and median values are calculated. Also the annotated treatment types are noted (“analgesia”, “sedation”, etc) so that an understanding of what treatments are administered and in which centres can be built into a definitive list. The treatment target is noted to match the treatment to the correct physiological event (i.e. only a treatment with a target of “CPP” is counted in response to a CPP physiological event).

It is noted here that in the implementation of this work (using the Brain-IT data-set – see introduction to section 5), the criteria are run against a total of 262 patients. The maximum number of analysis runs therefore becomes 33536 (128 * 262). It is beyond the scope of this current paper, but this points to further work required in the understanding of what treatments are administered and in which centres can be built into a definitive list.

4.2 Algorithms

The algorithms that drive this method are detailed in this sub-section. The output of the full process requires careful analysis as the structure of the association data is listed per patient, however the number count totals will be compiled as the code traverses the 262 patients (e.g. we want to know the total number of events that have associated treatments for ICP > 20 mmHg, with a hold-down of 10 minutes, within a time-window of 60 minutes. This needs to be totalled up from each patient, then distributed throughout the final totals).

4.2.1 Event-detection algorithm

To detect events from a physiological output stream, the following algorithm is used.

1) Compile the list of parameter objects ahead of processing. This is a list of the eight different definitions of ICP and CPP. A parameter in this context represents a physiological data stream – the parameter referring to a physical measurement of the patient’s brain. A representative parameter object is shown in figure 2.

![Image](image_url)

**Figure 2: ICU parameter object with the values required to define an event within the data stream (example values in brackets)**

This list of parameter objects constitutes part of the minimum required domain knowledge to allow event detection in a physiological data stream to occur. This can be read in from any persistent data store, such as a database, a properties file or an XML ontology file.

2) Querying the patient database: for each patient:

- Read the patient data into an “n x n” vector of vectors (i.e. a matrix).
- Each line in the object is a time-point (as the sampling rate is minute by minute, therefore each line increments by a minute) and each column is a particular parameter feed.
- The header line is used to identify the column index for the parameter that is of particular interest (ICPm, CPPm, etc).

3) For each parameter in the list compiled in step (1):

- Retrieve all of the parameter information for that indexed parameter object (name, unit, threshold, etc).
- For each hold-down definition:
  - Read in the line, timestamp and value (from step 2) and check the time between this timestamp and the last
    - if (gap > 1min)
    - Reset all event metrics and jump to end of the entire checking loop
    - if (event is in progress)
    - Is value still above threshold?
    - if (no)
      - Is the clear condition met?
      - if (yes)
if (potentialClear option is false)
    • Set the potentialClear variable to true
    • Increment the clear hold-down count

if (clear hold-down count equals the hold-down definition)
    • Note the event end time and add to event object
    • Add the event to the list of events
    • Increment the event index
    • Add value and timestamp to the event list

if (event is not in progress)
    • Is value still below threshold?
    • if (no)
        • Is the event condition met?
            • if (yes)
                o if (potentialEvent option is false)
                    • Set potentialEvent to true
                    • Set event hold-down count to zero
                o else
                    • Increment the event hold-down count
            • if (event hold-down count equals hold-down definition)
                • Note the event start time and add to event object
            • if (event condition NOT met)
                o Reset potentialEvent to false
                o Reset event hold-down count

To associate the events and treatments for each patient, all treatment information is retrieved, then for each parameter (defined in step (1) of event detection), for each hold-down definition, and for each time-window definition, the following association algorithm is run.

For each event:
    • Get the event start time
    • Define a time-window instance that begins at the event start and lasts for e.g. 30 minutes
    • For each treatment:
        o Get the treatment time
        o Get the treatment target, description and value
        o Isolate all treatment instances that have the tags “cpp”, “icp” or “hypotension” anywhere in the three string values
        o Isolate all the treatments that are end tags
        o If the treatment time is within the time-window bounds:
            • If the treatment is not an end tag and the event does not already have an associated treatment:
                • Add the treatment to list of treatments associated with this event
                • Set the Boolean flag indicating the event now has an associated treatment
                • Get the time to this treatment
                • Add the list of associated treatments to the time-window object for this event
                • If the associated treatment list is greater than zero:
                    o Increment the associated event counter
                • Add all time and treatment data gathered to the patient’s association data object and return this to the calling function.

Using this algorithm, the events are extracted for all 32 definitions.

4.2.2 Event/treatment association algorithm
For each patient an association object is instantiated, shown in figure 3.

Figure 3: a patient association object containing identifiers, a list of associated treatment times, treatment values (e.g. sedation etc), events and association numbers. A tree-map structure is used to store the indexed information so that the data can be retrieved in order when traversed for output.

The patient’s association data is then used to output the details to a text file and illustrative charts. As the association data is per-patient, the information must still be traversed for all summary counts and centre-specific information to be output.

These algorithms and processes form the core of the key functions required by the work so far. In the next section, the results achieved when applied to the real-world Brain-IT traumatic brain injury data-set are shown.

5 Results

The Brain-IT core data-set is a repository of 262 patients drawn from specialist neurological centres around Europe, collected with a view to enabling follow-on post-hoc analyses. One of the most comprehensive collections of high resolution TBI data with treatment and surgery
announcements to date, it forms a detailed retrospective view of physiological and treatment data that is well suited to analyses such as the one described in this paper. For more information about the specific composition of this dataset see Piper et al 2010.

Although the BrainIT dataset contains a high number of neurological ICU parameters (e.g. surgery, neurological response, demographics, etc) the work described in this paper focuses on the high-resolution physiological data and the annotated treatment data.

5.1 Data coverage
Preparatory to the analysis, the coverage of physiological parameters in the database is summarised in table 1. Coverage is defined by dividing the number of data points that are not “null” or blank by the overall number of data points for that physiological stream, and calculating the resulting percentage. Parameters with coverage less than 10% are omitted as contributing negligible information.

<table>
<thead>
<tr>
<th>ICU Parameter</th>
<th>Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>RR</td>
<td>26%</td>
</tr>
<tr>
<td>HRT</td>
<td>87%</td>
</tr>
<tr>
<td>BPs</td>
<td>84%</td>
</tr>
<tr>
<td>Bpd</td>
<td>84%</td>
</tr>
<tr>
<td>Bpm</td>
<td>96%</td>
</tr>
<tr>
<td>ICPm</td>
<td>84%</td>
</tr>
<tr>
<td>CVPm</td>
<td>20%</td>
</tr>
<tr>
<td>CPP</td>
<td>82%</td>
</tr>
<tr>
<td>TC</td>
<td>70%</td>
</tr>
<tr>
<td>SaO2</td>
<td>82%</td>
</tr>
<tr>
<td>SaO2pls</td>
<td>23%</td>
</tr>
<tr>
<td>ETCO2</td>
<td>19%</td>
</tr>
</tbody>
</table>

Table 1: physiological parameter coverage in the Brain-IT database

By inspecting the coverage for the data points used, a level of initial confidence can be gained to see how well represented the data-points are. If the parameter stream is well covered, it is a reasonable expectation that the event detection algorithm will produce useful information.

From the results shown in table 1, we can see that the blood pressure measures are all at least above 80%, including those of particular interest: mean intra-cranial pressure (ICPm) and cerebral perfusion pressure (CPP).

5.2 Event detection
The event detection algorithm is run across the 32 definitions referred to in section 4.1. Figure 4 shows the relevant results for ICP.

From figure 4 we can see that the most populous number of events in ICP come from definition #2 (a crossing threshold of > 15 mmHg). This represents a minima inflection point that will inform the next steps of treatment association. It is believed that definition #3 (< 70 mmHg) of CPP also represents this inflection point, but analysis time ran out before the writing of this publication and must be investigated further.

5.3 Treatment association
The overall number of treatments annotated in the dataset is 19175. The association of a treatment with an event provides a modifying parameter to the overall event number counts discussed in section 5.2. The inference...
that can be made is that general clinical definitions of events dictate that an ICP event occurs when a patient’s ICP crosses a threshold of 20 mmHg. The effects of this modification can be most clearly seen in the bar chart that represents the number of events with an associated treatment per definition per hold-down value, with a time-window of 30 minutes (time window #1), shown in figure 5.

The graph shape in figure 5 is evidently different from the event count numbers in isolation in figure 4. It is now definition #3 (> 20 mmHg) of ICP with hold-down of 5 minutes that appears to be most numerous, which would suggest that this definition input is triggering a larger number of clinical responses in terms of administered treatments. Again, the intuitive physical interpretation of these graphs can be seen in the shape of the distribution as the time-window increases towards an asymptote of infinite time. According to the association algorithm presented in section 4, the number of events with associations will approach the total as the time-window increases towards infinity (with a sufficiently large time-window, every event will have an associated treatment). Figure 6 (at the end of paper) demonstrates this progression of shape of the ICP events with treatment associations as they move through the other three time-window definitions (gradually definition #2 predominates again).

5.4 Treatment composition

For every combination of parameter definition, hold-down and association time-window, a composition of the actual treatments included in the list can be constructed. It is this information that will eventually inform the construction of a predominating treatment protocol. Figure 7 shows the treatment distribution for ICP definition #3 (> 20 mmHg) with a hold-down value of 5 minutes and a time-window for association of 30 minutes (the predominating EUSIG definition of an ICP event and “most reasonable” association time – see section 2).

The top three treatments applied in this instance are paralysis (18.2%), sedation (17.2%) and osmotic therapy (16.2%), from an overall number of 582 events with treatment associations (7.7% of the total event number for this ICP definition).

5.5 Centre-specific information

Using the unique centre reference identifier, event counts, associated treatment counts and treatment composition and times can be applied for each individual centre. For the purposes of discussing the technological ability to derive a centre-specific protocol, the results for the centre in Uppsala, Sweden are as follows:

The top three treatments are paralysis (32.9%), analgesia (13.4%) and a joint third place (11.4%) for ventilation, volume expansion and sedation, from an overall number of 373 events with treatment associations (3.1% of the total event numbers for this centre and ICP definition).

Combined with the metrics for treatment times (measured as well but not presented in this paper), a profile can be built up for a specific centre detailing frequency, response time, and how this profile compares to the guideline-mandated responses or specific study protocols.

6 Discussion and development

The results presented in section 5 describe the first steps in drawing physiological event and treatment information out of high resolution ICU data and using this to form an empirical treatment protocol. In this example we have been able to extract from the results a predominantly accepted clinical definition of an ICP event (> 20 mmHg, 5 minute hold-down).

There are several limitations in this analysis. One is that causality and association of two objects is a difficult process to accomplish. A major assumption has been the one-to-one relationship between an event and an annotated treatment. An estimated measurement of many-to-one (treatment to event) instances has been made and seems to be negligible but requires more exhaustive investigation.

Other limitations are that the structural information used to define a physiological monitoring event object may require modification. The structural description of a single event is well described in this paper, however as mentioned previously, a further extension of this definition, is to describe multiple events as a more complex structure. A collection of events within a specified period, with a specified gap between them, would indicate an “episode”, which would have further characteristics likely to be different in nature to an isolated event (for instance, does the notion of a hold-down time still apply to a collection of events?) Other types of events not captured in this work – such as very long-term “events” (e.g. a patient who has ICP just over the threshold for many days) or refractory events (those that end in patient death) – need to be considered. It should also be noted that the data-set used has been compiled by specialist centres that are known to focus on ICP and CPP measurements, so the detail in these readings may be unusually high.

Despite these limitations the initial results do show promise: - an accepted clinical definition has appeared...
and a usable treatment pattern has been derived, which, after further validation, can now be expressed in a workflow language to progress this research.

The next steps in validation are to run the same processes against different data-sets. There are several that have been identified as potentially useful in this space. First is the MIMIC II data-set at MIT (Saeed, Singh and Sanyal 2002). This is a publicly accessible collection of high-resolution ICU data. It is ideal in that it captures similar data-points to that of the Brain-IT dataset but is not specifically targeted to traumatic brain injury. If similar findings in ICP and CPP can be found then this would be a highly significant validation of the event detection and treatment association processes.

Second is the ICM+ (Smielewski et al 2005) data-stream based at the neuro-trauma unit at the Alfred Hospital in Melbourne, Australia. This is very high resolution data (waveform) that is targeted towards traumatic brain injury. The challenge, and independent validation, here would be condensing the information down to the same sampling rate (minute-by-minute) and understand if the processes produce the same result.

The third major data-set sourced to date that would be applicable, is the next generation of the Brain-IT data: the Avert-IT data repository (Stell et al 2009) with data capture tools (Philips CareView and Rhapsody products) which are in place at the Southern General Hospital in Glasgow, UK.

The question of data quality should also be addressed here – the treatment annotations of the Brain-IT dataset have been noted to not have a high degree of (temporal) accuracy. This was largely due to the manual methods of treatment information input that were required. As this is such an important issue in establishing the validity of association, these other data-sets have been specifically inspected to make sure the accuracy of the treatment time-points are as high as possible. Examples of this include the connection of intubation pumps to the bedside monitoring system (allowing the immediate recording of an automatic or manual drug administration) and touch-sensitive detection of hypothermic induction blankets, connected to the Philips Rhapsody system.

As patient information capture technology improves, it is assumed to be the case that the data quality and accuracy improves commensurately. However, another possible avenue to this work is to attempt to infer where a treatment has been administered by analysing the physiological output in a way opposite to the methods described in this paper. The potential value of “repairing” a patchy data-set such as this could be very high indeed, however the number of variables and uncertainty in this procedure would also be high. Either way, these two parallel strands (better capture technology and data improvement algorithms) will likely result in the same outcome: high quality data that provides more useful insight on the state of a patient’s physiology, which can contribute to the overall goal of improving study data. The numbers of associated events shown in section 5 are of low representation (7.7% overall and 3.1% for the specific centre), so all work to improve this would be wise to investigate and pursue.

If these validation procedures give rise to results similar to those described in this paper, it will be a significantly general technological solution to extracting treatment protocol information from widely available ICU data. However, in pursuit of the more ambitious goals of this overall work, the next step is to translate the treatment protocols derived into a form of abstract expression (referred to in the introduction as an object). A variety of possibilities exist in this space – for instance ProForma (Sutton and Fox 2003) is a clinical work-flow project that expresses guidelines as actions, conditions and states. But the primary characteristics that must be captured are a more detailed exploration of the nature of the treatment and event objects, and their individual and aggregate relationships. The key relationship measures will be the medical impact of the treatment, the temporal distance and causal characteristics between the treatment and event. In compiling a more complex protocol object that encapsulates this relationship, a key mathematical requirement will be the ability to parameterise such an object and allow it to be compared against others of a similar nature.

The ultimate goal of the work remains to analyse the data variation itself and find a way to compare this against other sets of data (either for trials and studies or drawn from general ICU data collection). The steps described in this paper are a successful attempt to draw out the initial information but many more challenges lie ahead to make this approach to standardising trial data a reality.

7 References


Figure 6 (left): the three other time-window definitions for ICP event and treatment association. As the time-window increases (vertically downwards), the distribution shape reverts back to that of the event count without treatment association (note the transposition of the two largest columns in particular). These provide the transition from figure 4 to figure 5.
Abstract

Introduction: The Caring Does Matter (CDM) programme aims to improve medication adherence amongst Pacific patients with high cardiovascular disease (CVD) risk. This paper examines CDM baseline (pre-intervention) data for patterns of poor medication adherence. Methods: Electronic medical records from 14 general practices are analysed with respect to prescribing patterns for antihypertensive, cholesterol lowering and oral antidiabetic medications. Patients who recently started treatment and had <80% medication possession in the latest 15 months are grouped into three categories: (1) just one prescription; (2) initial persistence (≤30 day lapse between the first and second prescriptions; and (3) other (i.e. multiple lapses, including a lapse immediately after the first prescription). Results: Over half of patients recently started on CVD-related medication are non-adherent in the latest 15 months; and the rate is higher than in those patients having started the medications earlier. A lapse after the first prescription is associated with significantly increased odds of non-adherence. Non-adherence is not dominated by a single prescription pattern category. Discussion: General practices usually get return visits from patients developing a pattern of poor medication possession, providing a series of signals of non-adherence risk, and offering ample opportunity for adherence promotion intervention.

Keywords: Cardiovascular risk management, electronic medical records, medication compliance, pharmaco-epidemiology.

1 Introduction

Poor adherence (also known as compliance) to long-term medication is a major issue undermining the effective delivery of healthcare (Rodgers and Ruffin, 1998). It is frequently overlooked by prescribing physicians when intensifying treatment (Heisler et al., 2008, Pittman et al., 2012). Statins, as a case in point, are a central element in cardiovascular disease (CVD) risk management as per guidelines in New Zealand (New Zealand Guidelines Group, 2012), Australia (National Vascular Disease Prevention Alliance, 2012) and internationally (Perk et al., 2012). The rate of failure to maintain statin therapy for 12 months after initiation is high (Benner et al., 2002) even when initiated after acute coronary events (Thornley et al., 2012). And poorer levels of statin adherence are associated with higher rates of long-term mortality after acute myocardial infarction (Rasmussen et al., 2007) and in coronary artery disease generally (Ho et al., 2008). New Zealand CVD guidelines place particular emphasis on the role of estimated 5-year risk of a cardiovascular event (e.g. heart attack and stroke) as central in the decision to treat – by prescribing statins or other relevant classes of medication, as well as through lifestyle modifications such as smoking cessation (New Zealand Guidelines Group, 2012).

EMRs related to medication supply (i.e. prescribing and dispensing) enable systematic estimation of medication adherence by indicating the availability of prescription medications to patients. One powerful statistic computable from EMRs is medication possession ratio (MPR), which is a percentage of days covered with
medication supply in some evaluation period; an MPR < 80% is commonly interpreted as indicating non-adherence (i.e. the EMRs showed that the patient would have lacked adequate supply of the medication at least one day in five). EMRs have been successfully used in New Zealand general practices to identify potential intervention targets with poor blood pressure (BP) medication adherence (Mabotuwana et al., 2009b) as well as to identify recipients for review who were on unchanged therapy while CVD risk and systolic BP remained high (Patel et al., 2013).

There are 266,000 Pacific people living in New Zealand (NZ), according to the 2006 census data (Statistics New Zealand and Ministry of Pacific Island Affairs, 2010). This population group have higher risk for CVD and higher mortality rate from CVD than the overall New Zealand population (Ministry of Health, 2012), but low adherence to CVD medications (Warren et al., 2012b). The Caring Does Matter (CDM) programme aims to improve Pacific people’s adherence to CVD medications by delivering structured primary care to the patients with high CVD risk (5-year event risk ≥ 10%) and low medication possession ratio (MPR < 80%) (Warren et al., 2012a). The CDM programme uses the general practice EMR to identify gaps in CVD medication supply (indicating poor medication adherence) in Pacific adults with high CVD risk.

This paper examines the CDM baseline data to understand the patterns of poor adherence for CVD medications by examining the medication supply in EMR prescribing records among the Pacific patients who became non-adherent to these medications. The objective is to gain insight into how medication non-adherence presents over time in the EMR data to better inform future interventions aimed at reduction of CVD event risk.

2 Methods

The CDM protocol was approved by the Northern X Regional Ethics Committee (NTX/12/EXP/102). The CDM baseline (pre-intervention) data from 14 CDM-participating general practices that use the MedTech EMR system (12 in Auckland and 2 in Northland) were analysed for the present study. This baseline data was extracted between May and September 2012. Data collected was de-identified prior to removal from the practice and included: ethnicity codes (up to three), age, gender, enrolment status and date enrolment commenced, and prescriptions for the previous five years. Three broad classes of CVD-related medication prescriptions – antihypertensive, cholesterol lowering and oral antidiabetic medications – are examined using the prescription records in the EMR by the SAS statistical software package, version 9.2 (SAS Institute Inc., Cary, North Carolina). These three medication classes are central to the CDM programme as each control one of the key risk factors known from epidemiological studies to increase risk of CVD events (Ho et al., 1993, Mannan et al., 2013):

1. Antihypertensive agents refer to the class of drugs used in the treatment of acute or chronic vascular hypertension (high blood pressure), including diuretics, adrenergic beta-antagonists, angiotensin-converting enzyme inhibitors and calcium channel blockers (National Library of Medicine, 2011) – herein treatments for chronic hypertension are the ones of interest.

2. Cholesterol lowering medications include statins and fibrates, as well as other drugs to manage hyperlipidemia (high cholesterol), such as ezetimibe (Pahan, 2006).

3. Oral antidiabetic medications lower the blood glucose level in patients with type 2 diabetes mellitus; these oral hypoglycaemic agents include sulfonylureas, meglitinides and biguanides (Luna and Feinglos, 2001).

Our primary interest in the present study is to characterise the development of poor adherence from the commencement of therapy in the above three classes of medication. To identify Pacific adults starting therapy, we apply the following inclusion criteria:

1. Self-identified as Pacific: any of the three ethnicity fields in the patient’s EMR is identified as Cook Island Maori, Fijian, Niuean, Other Pacific Island, Pacific Islander (Not Further Defined), Samoan, Tokelauan or Tongan.

2. Aged 20 or over at baseline.

3. Currently enrolled (at baseline) in the general practice and have been enrolled for at least three years.

4. In the last five years, the first prescription (for the class of medication being analysed, e.g. antihypertensives) occurred in one of the first three quarters of past two years (the ‘Run-in Period’ in Figure 1). In other words, the patient has no prescription record in the practice EMR for 3 years prior to the 9-month window of starting treatment.

5. In the last 15 months (the remaining five quarters of past two years, or ‘Evaluation Period’), the EMR recorded prescriptions on three or less distinct days, which signals the patient as being significantly undersupplied and having an MPR < 80%. (Given the usual NZ practice of prescribing a 90-day supply for long-term medications, five prescriptions for five quarters of therapy are expected; manual analysis confirms that virtually all prescriptions in these medication classes are for 90 days’ supply – e.g. a bottle of 30 with instructions to take one per day and with 2 refills.)

For each class of the CVD medication, we compute the inclusion eligibility and MPR with a SAS algorithm based on our previous work with the ChronoMedIt architecture (Mabotuwana and Warren, 2010). MPR < 80% in the Evaluation Period (most recent five quarters) is interpreted as non-adherence (and conversely MPR ≥ 80% is termed ‘high’ adherence). Note that this approach to identification of non-adherence is conservative – e.g. a patient might not choose to get a prescription dispensed, and they might not choose to take the dispensed medication; but without a prescription they are unlikely to
be in supply (with some exceptions, e.g. use of a family member’s medication). It is also tolerant, or at least conservative, with respect to changes of dose, changes of medication sub-class due to side-effects and combination therapy (e.g. multiple classes of antihypertensive prescribed concurrently) – in each case, additional prescriptions may cause us to overestimate supply but are unlikely to result in an underestimate.

If there are 120 or more days between two adjacent prescribing dates (essentially >30 day out of supply), we define this out-of-supply period as a ‘lapse.’ This duration of lapse is concerning even taking into account mild stockpiling of medication and transient events such as a brief hospital stay. The calculation of lapses allows us to group the non-adherers into three categories:

1. Just one prescription (i.e. a single prescription in the Run-in with no further prescription in that class during the Run-in or Evaluation Period);
2. Initial persistence (i.e. the first prescription had less than a 30-day lapse before the next prescription);
3. Other (i.e. multiple lapses, including a lapse after the first prescription).

The definition of these three categories is consistent with patterns of non-adherence in antidepressant therapy analysed previously with similar methods (Mabotuwana et al., 2011). Descriptive analysis of the three groups is analysed previously with similar methods (Mabotuwana et al., 2011). Descriptive analysis of the three groups is performed to test if these ‘not enrolled’ patients (e.g. historically enrolled elsewhere). This is to test the assumption that most patients get their long-term medication prescriptions for those who started treatment during and before the study Run-in Period.

As per Table 1, 28% (1627 out of 5744) of Pacific adults have had at least one antihypertensive prescription in the last five years; 22% for cholesterol lowering medication, and 15% for oral antidiabetic medications. Due to the high percentage of Pacific adults present in the practice EMRs not enrolled with the practice at baseline (21,042, 67%), we also examined the prescription records of these ‘not enrolled’ patients (e.g. historically enrolled then moved away or currently visiting as ‘casual’). Some of these patients have been prescribed CVD medications from the practices in the last five years. However, the percentage of these ‘not enrolled’ Pacific adult patients having CVD prescriptions in the last five years is only 3% for antihypertensives, 2% for cholesterol lowering medication and 1% for oral antidiabetic medications.

<table>
<thead>
<tr>
<th>Drug class</th>
<th>Having prescription in last five years</th>
<th>Starting treatment during Run-in Period</th>
<th>Starting treatment before Run-in Period</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antihypertensive medication</td>
<td>1627</td>
<td>109</td>
<td>1382</td>
</tr>
<tr>
<td>Cholesterol medication</td>
<td>1285</td>
<td>98</td>
<td>1099</td>
</tr>
<tr>
<td>Oral antidiabetic medication</td>
<td>841</td>
<td>60</td>
<td>691</td>
</tr>
</tbody>
</table>

Table 1: CVD Medications Prescribed to the 5,744 Pacific Adults Enrolled for at least 3 Years

3 Results

3.1 Study Participants

The EMR at 14 CDM-participating general practices recorded information for 49,088 Pacific patients, including 31,227 aged 20 or over. Among these Pacific adults, 10,185 (33%) were enrolled and funded at the practices at baseline, including 5,744 (18%) enrolled with the practice for three or more years. Considering only these 5,744 patients, Error! Reference source not found. shows their rates of having CVD medication prescriptions in the last five years, with separate tallies for those who started treatment during and before the study Run-in Period.
3.2 High Adherence levels

Among all the Pacific adults who fit the study inclusion criteria 1-4 (i.e. enrolled for at least 3 years and with the first prescription in the medication class starting in the Run-in Period), there are more non-adherers (i.e. having <80% MPR in the recent 15 months) than high adherers (≥80% MPR). Those Pacific adults starting treatment prior to the recent two years have significantly higher rates of adherence than those starting in the Run-in Period for all three medication classes (binomial proportion equivalence test, p<0.001); see Table 2.

To investigate the pattern of how a patient became a high adherer or non-adherer, we examined the MPR level of those meeting the study inclusion criteria 1-4 and if they persisted with the treatment initially (i.e. the first prescription had less than a 30-day lapse before the next prescription). Table 3 shows that those who persisted initially are more likely to adhere to their medications as compared to those who lapsed on first prescription (including those having only one prescription and multiple lapses). The odds ratio for high adherence for initial persistence status compared to initial lapse status is 8.05 for antihypertensives, 4.17 for cholesterol and 2.11 for antidiabetic medications, indicating increasing odds of high adherence for initial persistence. The 95% confidence interval of the OR indicates that the odds of high adherence for initial persistence are significantly higher for the oral antidiabetic medication class.

3.3 Categories of Non-adherers

The medication supply gaps identified in the non-adherers falls into three categories: (1) with just one prescription, (2) having persisted initially, or (3) having had multiple lapses (including a lapse after the first prescription). Table 4 shows that none of the three categories dominates the non-adherence pattern, except for a low rate of just-one-prescription cases for the oral antidiabetic medication class. However, it must be noted that in this fine-grained analysis the sample sizes are becoming quite small.

Figure 2 illustrates the initial persistence type of non-adherence pattern over two years using a case where a patient started Gliclazide (a sulfonylurea) and Metformin Hydrochloride (a biguanides) therapy in the first quarter of the two years. The timeline demonstrates some persistence initially but the lapses between prescriptions are tending to increase over time. However, at the CDM baseline data extraction time, the patient seems to be in supply, but only for Metformin Hydrochloride. The coloured areas indicate the 90-day coverage of each prescription from the date of prescribing.

Figure 3 shows an example of the multiple-lapse type of non-adherence pattern over two years for Simvastatin (a statin). The timeline view on prescription events demonstrates the first >30 day lapse right after the initial prescription as well as significant lapse beginning in Quarter 5, and a final lapse (continuing at the time of CDM baseline).

<table>
<thead>
<tr>
<th>Drug class</th>
<th>High adherers among those starting treatment during Run-in Period: N (% CI)</th>
<th>High adherers among those starting treatment before Run-in Period: N (% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antihypertensive medication</td>
<td>33 (30%, 22%-40%)</td>
<td>759 (55%, 52%-58%)</td>
</tr>
<tr>
<td>Cholesterol medication</td>
<td>27 (28%, 19%-38%)</td>
<td>441 (40%, 37%-43%)</td>
</tr>
<tr>
<td>Oral antidiabetic medication</td>
<td>20 (33%, 22%-47%)</td>
<td>339 (49%, 45%-53%)</td>
</tr>
</tbody>
</table>

Table 2: Adherence Rate of Enrolled Pacific Adults who Started Therapy during and before the Run-in Period

<table>
<thead>
<tr>
<th>Drug class (number initiating therapy in Run-in)</th>
<th>Persisted or Lapsed Initially</th>
<th>High adherers</th>
<th>Non-adherers</th>
<th>High adherence OR for ‘initial persistence’ compared to ‘initial lapse’ (CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antihypertensive medication (109)</td>
<td>Persisted initially (50)</td>
<td>26</td>
<td>24</td>
<td>8.05 (3.07, 21.12)</td>
</tr>
<tr>
<td></td>
<td>Lapsed initially (59)</td>
<td>7</td>
<td>52</td>
<td></td>
</tr>
<tr>
<td>Cholesterol medication (98)</td>
<td>Persisted initially (41)</td>
<td>18</td>
<td>23</td>
<td>4.17 (1.63, 10.71)</td>
</tr>
<tr>
<td></td>
<td>Lapsed initially (57)</td>
<td>9</td>
<td>48</td>
<td></td>
</tr>
<tr>
<td>Oral antidiabetic medication (50)</td>
<td>Persisted initially (35)</td>
<td>14</td>
<td>21</td>
<td>2.11 (0.68, 6.60)</td>
</tr>
<tr>
<td></td>
<td>Lapsed initially (25)</td>
<td>6</td>
<td>19</td>
<td></td>
</tr>
</tbody>
</table>

Table 3: Adherence Status by Persisted or Lapsed Initially

<table>
<thead>
<tr>
<th>Drug class</th>
<th>Just one prescription</th>
<th>Initial persistence</th>
<th>Multiple lapses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antihypertensive medication (76)</td>
<td>25 (33%)</td>
<td>24 (32%)</td>
<td>27 (36%)</td>
</tr>
<tr>
<td>Cholesterol medication (71)</td>
<td>19 (27%)</td>
<td>23 (32%)</td>
<td>29 (41%)</td>
</tr>
<tr>
<td>Oral antidiabetic medication (40)</td>
<td>3 (8%)</td>
<td>21 (53%)</td>
<td>16 (40%)</td>
</tr>
</tbody>
</table>

Table 4: Patient Number (%) in Non-adherence Categories
3.4 Characteristics of non-adherers

By definition, those in the just-one-prescription category had one lapse continuing through to the CDM baseline date. But between those who persisted initially (<30-day lapse between the first and second prescriptions) and those who did not persist but had multiple lapses for antihypertensive and cholesterol lowering medications, some differences are observed in terms of the mean number of lapses per patient over two years (see Table 5). Note that statistical significance was not tested given the small sample size.

Table 4 illustrates that patients who have come to have a low MPR often do so through multiple lapses – not just going away never to return. Table 6 provides rate and duration of final lapses (cases where the patient has not had a prescription for medication in the given class within 120 days of the end of the Evaluation Period). This shows that, among the non-adherers, a larger proportion of those who persisted initially than of those who had multiple lapses show long final lapse, and that those final lapses are longer than for the cases with multiple lapses that include a lapse immediately after the first prescription. As with Table 5, we refrain from testing statistical significance due to the small sample size.

4 Discussion

Analysis of primary care EMR data allows detailed examination of CVD medication supply and patterns of non-adherence. We found that rates of non-adherence after commencing therapy is high, with more than half of Pacific adults making an initial (as far as we could observe) start to CVD-related medication treatment having gaps in their medication supply consistent with non-adherence in the latter 15 months of treatment. The higher adherence level in those who have been on treatment for more than two years, as compared to those starting in the study Run-in Period, suggests that adherence levels may improve gradually over a period of several years. It could be that this is explained by the patients that started earlier having a greater burden of CVD-related morbidity; for instance, it is known that adherence rates are higher in those who have had a

<table>
<thead>
<tr>
<th>Drug class</th>
<th>Just one prescription</th>
<th>Initial persistence</th>
<th>Multiple lapses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antihypertensive medication</td>
<td>1</td>
<td>1.6</td>
<td>2.1</td>
</tr>
<tr>
<td>Cholesterol medication</td>
<td>1</td>
<td>1.7</td>
<td>2.1</td>
</tr>
<tr>
<td>Oral antidiabetic medication</td>
<td>1</td>
<td>1.9</td>
<td>1.9</td>
</tr>
</tbody>
</table>

Table 5: Mean Number of Lapses per Patient over Two Years among Non-adherers

<table>
<thead>
<tr>
<th>Drug class</th>
<th>Just one prescription</th>
<th>Initial persistence</th>
<th>Multiple lapses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antihypertensive medication</td>
<td>100%, 498</td>
<td>67%, 253</td>
<td>44%, 178</td>
</tr>
<tr>
<td>Cholesterol medication</td>
<td>100%, 519</td>
<td>65%, 264</td>
<td>45%, 180</td>
</tr>
<tr>
<td>Oral antidiabetic medication</td>
<td>100%, 440</td>
<td>48%, 268</td>
<td>19%, 176</td>
</tr>
</tbody>
</table>

Table 6: Non-adherers with Final Lapse by Category (% Mean Number of Days in Final Lapse)
myocardial infarction than those who have not (Naderi et al., 2012). Another possibility is that the patients that appear newly started by our inclusion criteria were ones who had lapsed for all of the year before the Run-in Period (and for up to two year before that if they had been with the practice that long), while not being genuinely new to the therapy. Indeed it seems probable that we are observing a mixture of these effects; the degree to which these factors are in play will require further analysis to assess. The prescription pattern for the patients with low medication possession demonstrates that the adherence gaps are not dominated by ‘just-one-script’ cases. The presence of subsequent prescriptions of the same class in the majority of low medication possession cases indicates: (a) that the practice GPs are persisting in the belief that therapy in the given medication class is appropriate for that patient (i.e. they have not re-assessed the CVD risk or concluded that it is contraindicated); and (b) that patients are continuing in their relationship with the practice and concur to the therapy insofar as to accept another script (but not sufficiently to achieve high adherence). This means that the practice typically gets the opportunity for ‘another look’ at patients heading into poor CVD medication adherence, and so has a chance to take some further responsibility and action to promote medication adherence. We observe that the odds of non-adherence in the longer term (the latter 15 months of the two years that include initiation of therapy) are significantly greater if patients lapse after the very first prescription – this should be taken by practices as a ‘red flag’ to trigger adherence promotion activity.

This study has a number of limitations. The set of practices is a convenience sample based on practices willing to participate in the CDM initiative. Moreover, our analysis is based on patients that were still enrolled with a practice after three years; in day-to-day prescribing, from a practice perspective, non-adherence will be more frequent due to patients that have changed enrolment to a different practice. However, if a practice had a policy of systematic follow-up for poor medication supply, such enrolment changes would be quickly revealed. Moreover, we found that rates of prescribing to ‘casual’ patients in the CVD-related medication classes analysed in this study were very low (only 1-3% of patients) – thus, a change in CVD medication prescriber appears to track well with a change in enrolment. Supply-based MPR is, of course, an indirect measure of adherence, although widely accepted due to its practical applicability at the population level as compared to direct monitoring, and with less vulnerability to over-estimating adherence as compared to pill counts or self-report (Andrade et al., 2006, Steiner and Prochazka, 1997, Vermeire et al., 2001). While we have based our adherence assessment only on prescriptions, we have found previously that general practice prescriptions for long-term medication match well with national reimbursement data for dispensing (Mabotuwana et al., 2009a) and that – at least for antihypertensive medication – improved MPR based on prescribing translates to improved MPR on dispensing (Warren et al., 2012b). Although we started with a large cohort of Pacific adults, tracking non-adherence for those starting long-term therapy in a narrow time period led to small sample size for some of the more fine-grained analysis. It may be that linkage to national data is the only method that can build a population model that is robust across enrolment changes; however, the richness of general practice EMRs should not be abandoned in the process.

While there is widespread agreement that medication adherence is a major problem in management of CVD risk (Lemstra et al., 2012), we believe that the temporal dimension of the phenomenon has been under-analysed. Our analysis indicates that a general practice usually makes repeated prescriptions (i.e. subsequent prescriptions in the same broad medication class) to patients developing a pattern of poor medication possession, providing a series of signals and ‘red flags’ of non-adherence risk, and offering ample opportunity for adherence promotion intervention.

5 Acknowledgments

This project was made possible through funding from the Pacific Grant Fund, a fund of the Ministry of Health. We would also like to thank the West Fono management and staff, especially Tevita Funaki and Anthony Tutiahi for supporting the programme. We extend our thanks to the staff of the participating general practices for accommodating the study.

6 References


Author Index

Birch, Kate E., 3
Borsboom, Alexander S., 13
Boyce, Tai, 89
Day, Karen, 51
de Vries, Denise, 59
Gray, Kathleen, iii, vii, 21
Gu, Yulong, 89
Heffernan, Kayla J., 3
Kennelly, John, 89
Lopez Campos, Guillermo Hugo, 31
Martin Sanchez, Fernando Jose, 21, 31
Merolli, Mark, 21
Moss, Laura, 79
Ofoghi, Bahadorreza, 31
Orr, Martin, 51
Ou, Ying, 41
Patrick, Jon, 41
Peni, Bernadette, 51
Perimal-Lewis, Lua, 59
Piper, Ian, 79
Riddle, Patricia, 69
Song, Shen, 69
Stell, Anthony, 79
Thompson, Campell H., 59
Verspoor, Karin, 31
Warren, Jim, iii, vii, 69, 89
Listed below are some of the latest volumes published in the ACS Series Conferences in Research and Practice in Information Technology. The full text of most papers (in either PDF or Postscript format) is available at the series website http://crpit.com.

Volume 124 - Database Technologies 2012

Volume 125 - Information Security 2012

Volume 126 - User Interfaces 2012
Edited by Haifeng Chen, Flinders University, Australia and Rose T. Smith, University of South Australia, Australia. January 2012. 978-1-921770-07-4. Contains the proceedings of the Thirteenth Australasian User Interface Conference (AUIC2012), Melbourne, Australia, 30 January – 3 February 2012.

Volume 127 - Parallel and Distributed Computing 2012
Edited by Junjun Chen, University of Technology, Sydney, Australia and Rajiv Ranjan, CSIRO ICT Centre, Australia. January 2012. 978-1-921770-08-1. Contains the proceedings of the Tenth Australasian Symposium on Parallel and Distributed Computing (AusPDC 2012), Melbourne, Australia, 30 January – 3 February 2012.

Volume 128 - Theory of Computing 2012

Volume 129 - Health Informatics and Knowledge Management 2012
Edited by Kerryn Butler-Henderson, Curtin University, Australia and Andrew Trotman, University of Otago, New Zealand. January 2012. 978-1-921770-11-1. Contains the proceedings of the Fifth Australasian Workshop on Health Informatics and Knowledge Management (HIKM 2012), Melbourne, Australia, 30 January – 3 February 2012.

Volume 130 - Conceptual Modelling 2012

Volume 133 - Australian System Safety Conference 2011

Volume 134 - Data Mining and Analytics 2012
Edited by Xuezheng Zhao, Department of Immigration and Citizenship, Australia, Jiuyong Li, University of South Australia, Australia and Peter Christen, Australian National University, Australia. December 2012. 978-1-921770-14-2. Contains the proceedings of the Tenth Australasian Data Mining Conference (AusDM’12), Sydney, Australia, 5-7 December 2012.

Volume 135 - Computer Science 2013

Volume 136 - Computing Education 2013

Volume 137 - Database Technologies 2013

Volume 138 - Information Security 2013

Volume 139 - User Interfaces 2013

Volume 140 - Parallel and Distributed Computing 2013

Volume 141 - Theory of Computing 2013

Volume 142 - Health Informatics and Knowledge Management 2013

Volume 143 - Conceptual Modelling 2013

Volume 144 - The Web 2013