Semantic Interoperability for Health Network of Excellence

Deliverable 2.1
Summarising Health Records for Populations: Cardiovascular Use-cases

Cardiovascular healthcare use-cases for semantically interoperable summaries of health records at the population level

[Version] V 1.1
[Date] 31st December 2012
Call: FP7-ICT-2011-7
Grant agreement for: Network of Excellence (NoE)
Project acronym: SemanticHealthNet
Project full title: Semantic Interoperability for Health Network
Grant agreement no.: 288408
Budget: 3.222.380 EURO
Funding: 2.945.364 EURO
Website: www.semantichealthnet.eu
Coordinators:

The SemanticHealthNet project is partially funded by the European Commission.
Deliverable: 2.1 Summarising Health Records for Populations (M12)

Publishable summary: This report outlines typical requirements for summarising health records at the population level, using examples in cardiovascular healthcare.

We consider five population level uses of health information for preventive and early healthcare: 1) public health and social interventions; 2) clinical audit and optimising healthcare services; 3) payer evidence and commissioning healthcare services; 4) consumer health applications; 5) research using health records.

We use a series of vignettes in early and preventive cardiovascular healthcare to illustrate the functions of population health information. Each vignette involves a decision regarding multiple patients or citizens, which needs supporting with information derived from a variety of health records or related data sources. The first decision concerns the level of cardiovascular risk at which statins should be offered for primary prevention (preventing or delaying the onset of disease). Second we consider targeting scarce public health resources for reducing overweight and obesity in society. The third example is about monitoring the quality of cardiovascular healthcare services to promote early intervention, thereby achieving secondary prevention (slowing the progression of disease). Finally we anticipate the value of patients co-producing their own health records, adding a new level of longitudinal information that could transform decision-making for prevention and treatment.

The report is structured as follows: 1) an introduction to preventing cardiovascular disease; 2) an introduction to modelling disease risks and treatment outcomes to inform public health and healthcare quality improvement decisions; 3) an overview of relevant health data sources and the emergence of ubiquitous linkable data; 4) vignettes to illustrate population level uses of healthcare records for early or preventive cardiovascular healthcare; and 5) appendices I-IV, providing some deeper case studies. Appendix I considers missed opportunities for blood pressure control in preventing myocardial infarction or stroke. Appendix II illustrates the incompleteness of current health records in reflecting the true picture of heart failure care. Appendix III introduces the Collaborative Online Care Pathway Investigation Tool (COCPIT) for comparing ideal care pathways with the care that is recorded in health records. Appendix IV explores the potential to use simulation tools in health policy making.
### Document History

<table>
<thead>
<tr>
<th>Date</th>
<th>Revision</th>
<th>Author(s)</th>
<th>Changes</th>
</tr>
</thead>
<tbody>
<tr>
<td>22 Nov 2012</td>
<td>First draft</td>
<td>David Lamb</td>
<td>Changed Appendix II</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Added Publishable Summary</td>
</tr>
<tr>
<td>23 Nov 2012</td>
<td>Second draft</td>
<td>David Lamb</td>
<td>Changes to 3.3 and Vignettes</td>
</tr>
<tr>
<td>10 Dec 2012</td>
<td>Third draft</td>
<td>David Lamb</td>
<td>Changes to 3.3 and Vignettes</td>
</tr>
<tr>
<td>11 Dec 2012</td>
<td>Fourth draft</td>
<td>David Lamb</td>
<td>Changes to Vignettes</td>
</tr>
<tr>
<td>14 Dec 2012</td>
<td>Fifth draft</td>
<td>Amendments by Christi Deaton and Simon Capewell</td>
<td>Amendments to text</td>
</tr>
<tr>
<td>21 Dec 2012</td>
<td>Sixth draft</td>
<td>David Lamb</td>
<td>Adapting vignettes</td>
</tr>
<tr>
<td>26 Dec 2012</td>
<td>Seventh draft</td>
<td>David Lamb</td>
<td>Revising introduction and amending vignettes</td>
</tr>
<tr>
<td>31 Dec 2012</td>
<td>Eighth draft</td>
<td>Iain Buchan</td>
<td>New summary, new sections and edits throughout</td>
</tr>
</tbody>
</table>
# Table of Contents

Table of Contents .................................................................................................................. 5  
Introduction .......................................................................................................................... 6  
   The SemanticHealthNet Project .......................................................................................... 6  
   Executive Summary ............................................................................................................ 8  
Information for Preventing Cardiovascular Disease ................................................................ 9  
   European Guidelines ......................................................................................................... 9  
   Clinical and Public Health Approaches to Prevention ......................................................... 11  
   Levels of Prevention ......................................................................................................... 12  
   Primary Prevention: Cardiovascular Example .................................................................... 12  
   Secondary Prevention: Cardiovascular Example ............................................................... 13  
   References ......................................................................................................................... 13  
Data Sources .......................................................................................................................... 13  
   Monitoring Risk Factors ................................................................................................. 13  
Vignettes: Population Summaries of Cardiovascular Health(care) ......................................... 15  
   Optimal Primary Prevention with Statins .......................................................................... 15  
   Targeting Public Health Interventions to Reduce Obesity .................................................. 18  
   Reconciling Methods of Monitoring Smoking Cessation .................................................... 20  
   Mapping Missed Opportunities to Prevent Cardiovascular Events .................................... 22  
   Monitoring the Diagnosis and Treatment of Heart Failure .............................................. 24  
   Patient Co-produced Information for Cardiovascular Health(care) ................................... 27  
   Real-world Trial of a New Medicine to Prevent Stroke ..................................................... 29  
Summary and Conclusions .................................................................................................... 31  
Appendix I: Missed Opportunities Mapping .......................................................................... 32  
Appendix II: Heart Failure in Primary Care ......................................................................... 39  
Appendix III: COCPIT ............................................................................................................ 43  
Appendix IV: Sharable Simulations of Health Policies ........................................................... 48
Introduction

The SemanticHealthNet Project

Semantic interoperability of EHR systems is a vital prerequisite for enabling patient-centred care and advanced clinical and biomedical research. SemanticHealthNet will develop a scalable and sustainable pan-European organisational and governance process to achieve this objective across healthcare systems and institutions.

A clinical focus on chronic heart failure and cardiovascular prevention in the workplan will drive the semantic resources to be developed. The exemplars in cardiology and public health are specific enough to permit comprehensive development and validation of these resources, and yet typical enough for wider generalisation of the methodology and its governance. SemanticHealthNet will capture the needs articulated by clinicians and public health experts for evidence-based, patient-centred integrated care in these domains. Existing European consensus in the management of chronic heart failure and cardiovascular prevention will then be integrated in EHR architectures, clinical data structures, terminologies and ontology by leading technical experts.

Clinical and Industrial Advisory Boards will provide links with other domains in which these results can be used beneficially. The project will investigate how best to combine and adapt informatics resources to support semantic interoperability, and how these can be developed and supported at scale. Results of this investigation will be generalised and formalised. The involvement of health authorities, clinical professionals, insurers, ministries of health, vendors, and purchasers will ensure that the project approach and results are realistically adoptable and viable. This work will also build on the SemanticHEALTH and CALLIOPE roadmaps for eHealth interoperability.

A business model to justify strategic investments, including the opportunity costs for key stakeholders such as Standards Development Organisations and industry, will be defined. Links with the epSOS large scale pilot and the eHealth Governance Initiative, will inform the shape of the Virtual Organisation that this Network will establish to sustain semantic interoperability developments and their adoption.

The consortium comprises 17 Partners and more than 40 internationally recognised experts, including from USA and Canada, ensuring a global impact.

Partners

1. Research in Advanced Medical Informatics and Telematics (RAMIT) – BE (Admin Coordinator)
2. Imperial College London (Imperial) – UK
3. University of Hull (UHULL) – UK
4. University Hospitals of Geneva (HUG) – CH
5. World Health Organization (WHO) – CH
6. The University of Manchester (UoM) – UK
7. Medical University of Graz (MUG) – AT
8. International Health Terminology Standards Development Organisation (IHTSDO) – DK
9. Institut National de la Santé et la Recherche Médicale (INSERM) – FR
10. Ocean Informatics (Ocean) – UK
11. Health Level 7 International Foundation (HL7 International) – BE
12. EN13606 Association (EN13606) – NL
13. Empirica Gesellschaft für Kommunikations- und Technologieforschung mbH (EMPIRICA) – DE
14. Standing Committee of European Doctors (CPME) – BE
15. European Coordination Committee of the Radiological, Electromedical and Healthcare IT Industry (COCIR) – BE
16. Whittington NHS Trust (WHIT) – UK
17. European Institute for Health Records (EuroRec) – FR (NoE Coordinator)

**Project Plan**

**Workstream I:**

- WP1: Patient care exemplar (heart failure)
- WP2: Public health exemplar (cardiovascular disease prevention)
- WP3: Stakeholder validation

**Workstream II:**

- WP4: Harmonised resources
- WP5: Infostructure and tools
- WP6: Industrial engagement

**Workstream III:**

- WP7: Adoption and sustainability
- WP8: European Virtual Organisation
- WP9: Project management, dissemination, promotion
Executive Summary

This report outlines typical requirements for summarising health records at the population level, using examples in cardiovascular healthcare.

We consider five population level uses of health information for preventive and early healthcare:

1) public health and social interventions;
2) clinical audit and optimising healthcare services;
3) payer evidence and commissioning healthcare services;
4) consumer health applications;
5) research using health records.

We use a series of vignettes in early and preventive cardiovascular healthcare to illustrate the functions of population health information. Each vignette involves a decision regarding multiple patients or citizens, which needs supporting with information derived from a variety of health records or related data sources. The first decision concerns the level of cardiovascular risk at which statins should be offered for primary prevention (preventing or delaying the onset of disease). Second we consider targeting scarce public health resources for reducing overweight and obesity in society. The third example is about monitoring the quality of cardiovascular healthcare services to promote early intervention, thereby achieving secondary prevention (slowing the progression of disease). Finally we anticipate the value of patients co-producing their own health records, adding a new level of longitudinal information that could transform decision-making for prevention and treatment.

The report is structured as follows:

1) an introduction to preventing cardiovascular disease;
2) an introduction to modelling disease risks and treatment outcomes to inform public health and healthcare quality improvement decisions;
3) an overview of relevant health data sources and the emergence of ubiquitous linkable data;
4) vignettes to illustrate population level uses of healthcare records for early or preventive cardiovascular healthcare; and
5) Appendices I-IV, providing some deeper case studies.

I) Appendix I considers missed opportunities for blood pressure control in preventing myocardial infarction or stroke.
II) Appendix II illustrates the incompleteness of current health records in reflecting the true picture of heart failure care.
III) Appendix III introduces the Collaborative Online Care Pathway Investigation Tool (COCPIT) for comparing ideal care pathways with the care recorded in health records.
IV) Appendix IV explores the potential to use simulation tools in health policy making.
Information for Preventing Cardiovascular Disease

Despite steadily declining rates across much of Europe, deaths from cardiovascular disease (CVD) still account for 42% of premature (age of 75) mortality among women and 38% among men. Furthermore, most of these early deaths are preventable.

In this report we demonstrate how information systems can be used to prevent early death and suffering from CVD. We demonstrate that efficient and imaginative use of data can underpin further reductions in the burden of CVD for society.

European Guidelines

There is a wealth of knowledge about the causes of CVD and preventive measures. However, there are gaps and inconsistencies in this knowledge base. In 2012, the Fifth Joint Task Force of the European Society of Cardiology (ESC) produced a summary of present knowledge in preventive cardiology. These guidelines are aimed at physicians and other health workers.

The fact that the ESC Guidelines are the fifth issued since 1994 reflects the rapid change in knowledge. Prevention of CVD is important because the condition is a major cause of early death and disability, and its treatment consumes a large proportion of national healthcare budgets. Thus there are major public health and economic gains to be made from preventing CVD – and there is a substantial body of evidence on effective measures for achieving this.

The ESC Guidelines are the product of agreement across the members of nine cognate societies. The report is structured around answers to five basic questions. The first concerns the nature of CVD prevention. The authors acknowledge that the distinction between primary \(^1\) and secondary \(^2\) prevention of CVD is artificial. The underlying disease process in blood vessels (atherosclerosis) is continuous, probably beginning in infancy.

Although the ESC literature search of relevant clinical guidelines identified more than 1900 publications, only seven of these were of a standard considered as showing ‘considerable rigour’. The guidelines present recommendations for practice, and the strength of recommendation uses the GRADE rating system, based on: 1) the level of evidence; 2) the risk-benefit ratio; 3) patient preferences; 4) and available resources.

The value of this approach is that if the evidence from a systematic review or clinical trial is biased, inconsistent or imprecise, the value of the evidence is downgraded. Similarly, observational data from cohort or case-control studies may be upgraded from low or moderate to high if it is unlikely that the evidence is biased, and the findings are consistent and precise. This has particular importance in relation to CVD prevention, where many of the interventions involve change in lifestyle, which are not easily evaluated in conventional trials.

---

\(^1\) Primary prevention is about preventing or delaying the onset of symptomatic disease.

\(^2\) Secondary prevention is about slowing the progression of diagnosed disease and preventing major events.
There is considerable evidence that more than 50% of the recent reduction in CVD is due to decreases in risk factors, including smoking. Despite this reduction there remains a significant proportion of the population who are exposed to high levels of modifiable risks. Two risk factors, obesity and diabetes, are rising substantially. As with smoking, these factors cause damage over many years, yet current practice focuses risk factor control on middle-aged or older people. There remains a major problem of intervening early when people have disease risks but no symptoms.

The ESC Guidelines consider methods for assessing individuals’ risks of suffering CVD events in a given number of years. The authors note that there is no threshold value of risk for initiating an intervention. Risk is a continuum so there is no exact point above which an intervention is indicated nor below which even the simplest measures, such as lifestyle advice, are not relevant.

The ESC note the importance of distinguishing relative and absolute risk. The risk charts illustrate how a young person with a low absolute risk may be at high but reducible relative risk of CVD. There is evidence that patients tend to underestimate their risk of CVD. The charts are an aid to the clinician in motivating patients to change lifestyle or adhere to prescribed medication. CVD risk tools are disseminated to aid decision-making, but it we note that risk estimation is more inaccurate than many patients/citizens might expect. The person with central obesity, a sedentary lifestyle or low HDL cholesterol may be at much greater risk of CVD than their score implies.

The largest section of the ESC Guidelines deals with how to use preventive measures. Two general approaches, lifestyle (e.g. smoking) and clinical (e.g. blood pressure) are presented. It is acknowledged that clinicians are ill equipped to support lifestyle change. We recognise the need for persuasive information that is part of daily life in order to achieve large-scale CVD prevention.

There is overwhelming evidence that stopping smoking, eating a healthy diet and taking regular physical exercise can control most CVD risk. However, there is a paucity of evidence of how to help citizens achieve this en masse.

Clinical interventions to control blood pressure, blood glucose and lipids, and the use of antithrombotic treatments, raise questions over targets, which patients to treat and the relative (cost) effectiveness of alternative drugs. The evidence base continues to develop, and “real-world” evidence, i.e. outside the artificial environments of clinical trials, will be crucial – and this requires large numbers of health records to be analysed.

Patient adherence to medication to prevent CVD is low. Evidence on the reasons for patient non-adherence is emerging but of variable quality.

The final section of the ESC report discusses the settings where prevention of CVD should be offered. Nurses, doctors and paramedical staff have a clear role in prevention. However since risk factors related to lifestyle are a recognised cause of CVD, conditions in the wider society have a major part to play. Having access to leisure facilities, a smoke free environment and high quality nutrition are all relevant to CVD prevention. Both clinical and political considerations matter.

**Key Informatics Issues:** The ESC report highlights the lack of knowledge about: i) people who do not usually participate in clinical trials; and ii) long term outcomes of interventions. This points to the
need for reusing health records to produce “real world evidence” for underpinning CVD policies. In addition, there is a need to reach beyond the clinic into citizens’ daily lives in order to understand how to modify CVD risk factors. When evidence is synthesised into guidelines the underlying risk estimation usually applies more to populations than to individuals, yet decisions about individuals are informed by these rule bases. The reliable modelling of individual vs. population CVD risk is a grand challenge that requires considerable Informatics effort.

**Clinical and Public Health Approaches to Prevention**

Here we distinguish clinical from public health interventions to prevent CVD.

The clinical approach titrates interventions to individual patient characteristics, whereas the public health approach considers the characteristics of groups of people.

The focus of clinical medicine is treatment of the individual, with the aim of improving the care and clinical outcomes of each individual. The population approach considers the characteristics of an aggregate of individuals, with the aim of improving their average health and wellbeing. Knowledge of the characteristics of a population can be derived not only from the aggregation of clinical records but from other sources such as surveys.

A basic assumption of the public health approach is that the health of a population is the result of the combined effect of many factors operating at different levels. At the highest level, the economic and social background of a population, such as occupation and education, are strongly associated with its morbidity and mortality rates. The physical environment, such as quality of housing, has a direct effect on health. Individuals’ genes interact with these other factors. Some factors can be changed so that the overall health of a group of people can be improved. For example reducing smoking and increasing levels of physical activity will deliver wide-ranging improvements in a population’s health.

Public health measures assume that: 1) the average health of a population is determined by the level of risk factors averaged across that population; and 2) if the average level of risk factors in the population is reduced then the burden of morbidity and mortality will also be reduced. For example, the public health service for a city may focus on smoking cessation because that city has higher smoking rates than the national average. The outcomes of public health measures will be measured in terms of process (i.e. numbers taking up a smoking cessation service), intermediate risk outcomes (e.g. six month quit rates) and disease outcomes (CVD event rates; CVD and lung cancer incidence and death rates).

The clinical approach is more opportunistic, and reactive. But at the overlap between primary healthcare and public health there are hybrid activities such as population screening for CVD among citizens in specific age groups. The following table draws the main distinctions between clinical and public health measures:-
Levels of Prevention

Prevention is defined in different levels: primordial (eradicating a risk factor like smallpox); primary (preventing or delaying the onset of a disease); secondary (slowing the progression of a disease after it has been diagnosed, in order to prevent adverse events); and tertiary (timely, effective treatment to lessen the suffering following disease events).

Primary Prevention: Cardiovascular Example

As an example consider the association between serum cholesterol and the incidence of CVD. People with an elevated level of cholesterol (5 mmol/l or more) are at increased risk of CVD compared to those with a level below 5 mmol/l. Many of those with raised cholesterol are unaware of their condition and the opportunity to lower their risk of CVD. Screening and offering medication to people with elevated cholesterol is an example of primary prevention. Reducing saturated fat in diet is also a way to reduce cholesterol, but all citizens can benefit from such lifestyle changes so this is a public health not just a clinical means of primary prevention.

Most public health measures in CVD have beneficial effects on other diseases too. For example, smoking cessation reducing lung cancer profoundly. Increasing physical activity and reducing obesity lowers the risks of many diseases and intermediate CVD risks such as blood pressure and cholesterol. An individual who quits smoking will increase their life expectancy: a 45 year old male smoker who quits will increase his life expectancy, on average, by 5.6 to 7.1 years (Taylor, 2002). It is not, however, realistic to forecast precisely the gain for any specific individual. This information is most accurate and potent in public health terms: for example, it has been estimated that reducing the prevalence of UK smoking by 1 percentage point each year for 10 years would prevent 69,049 premature deaths over that period (Lewis et al, 2005).
Secondary Prevention: Cardiovascular Example

People with a blood pressure of 140/90 mmHg or more are defined as hypertensive, and are at high risk of developing CVD. For individuals with hypertension, reducing blood pressure by 7.5 mmHg reduces the heart attack risk to 81% (or by 19%). While this can be difficult to explain to an individual patient the interpretation at the group level is simpler. If 100 people with an average blood pressure of 150/100 mmHg have the upper value of their blood pressure reduced by 7.5 mmHg there would be 19 fewer heart attacks among those people than would otherwise have occurred. It is unrealistic to forecast the health gain for any individual patient but the improvement for a population is predictable. Reducing the average blood pressure of a population will reliably reduce heart attacks.

References


Data Sources

The sources of data for monitoring CVD risks, interventions and outcomes include: medical records, social care records, personal health records, service payments, service quality information, demographic and public administrative data, area-level socio-economic data, health surveys, occupational health records, disease registers, vascular screening programmes and mobile health/wellbeing devices/applications.

Monitoring Risk Factors

Information on the lifestyle risk factors for CVD are not easily collected within a clinical setting. Self-reported measures, for example of smoking, alcohol consumption and physical activity, tend to under-report risks.

At community level large amounts of data are collected for purposes of administration, but they also reflect CVD risks. For example supermarkets, pharmacies and leisure facilities issue cards that record the transactions of customers. These datasets are used for marketing, via models of household characteristics and behaviours. Such models were first developed to inform advertising campaigns but have now become social marketing tools for other purposes. Their use in modelling lifestyle and health outcomes is emerging.

There are many occasions where data on risk factors can be collected opportunistically. Students attending further and higher education can be incentivised to give data on their health status or knowledge of health by completing on-line questionnaires and receiving e.g. a month’s free gym membership (e.g. www.advice.salford.ac.uk/page/healthy-eating). Data on young people is of particular value as they are not often seen by healthcare services.
Some risk factors, such as obesity, are the subject of national monitoring programmes such as the UK’s National Child Measurement Programme. Looking across the life course, childhood obesity is an important predictor of adult obesity at population level. Similar life course views of all of the major CVD are key to health policy making.

In many populations there is a willingness to provide personal health data where doing so can benefit society. For example around 400,000 people have provided samples, clinical measures and lifestyle information to the UK Biobank study. The participants agreed that the data stored in the research database can be linked to their healthcare and death records. Similar initiatives, such as the Born in Bradford birth cohort and Salford Citizen Science project, have been built around the local public health benefit from citizens participating in measuring health.

The consumer health and wellbeing sector is growing rapidly, and citizens are using mobile technologies to collect vast amounts of information relating to CVD. This information can be direct, for example blood pressure cuff and bathroom scales data transmitted to personal health records. It can also be indirect, for example physical activity inferred from position and motion sensing in mobile phones.

Large scale linked data of the future will be key to uncovering the true complexity of CVD risks.
Vignettes: Population Summaries of Cardiovascular Health(care)

The purpose of the following vignettes is to provide examples where population level summaries of individual health records can be employed to improve CVD prevention and early intervention.

Optimal Primary Prevention with Statins

Decision: At what level of CVD risk should statins be prescribed in my local health economy?

Background

There is a wealth of evidence to demonstrate that statins reduce deaths from CVD. Statins are also effective in preventing the development of CVD among individuals at high risk of a cardiovascular event within 10 years. Offering screening to a large population of people who do not have any apparent symptoms of CVD is not an effective use of resources. Thus screening is usually targeted at groups known to be at high risk. This vignette outlines an approach to targeting groups in the population who would most benefit from statins.

There are three common sources of intelligence for estimating the level of CVD risk in an asymptomatic population:

1. Scientific literature: There is a substantial body of research describing the effect of statin prescribing on CVD risk factors.
2. National surveys of the lifestyle and morbidities of the population.
3. Electronic health records (EHRs): These vary in completeness, measurement accuracy and coding. Aggregated EHRs are increasingly used to derive measures of population health. It is hoped that patients will provide more accurate lifestyle information into EHRs in future, either directly or via linked personal health records/apps.
4. Disease registers (increasingly derived from EHRs rather than supplementary data entry).
5. Death certificates and other public administrative data.

Data from these sources are used to model CVD risk. The risk is usually expressed in terms of: developing CVD (incidence); experiencing an event (e.g. heart attack); or dying – with a time horizon of 10 years.

The cost of a programme to screen people for CVD risk depends on several factors. The number of people invited is one cost, although not the major one. The asymptomatic patients at highest risk of CVD are usually the least likely to respond to an invitation to be screened, for example people from deprived areas. The cost of persuading ‘hard to reach’ groups to attend screening should be factored in.
Typical Data Inputs

Research findings quantifying the effects of statin prescribing on the incidence of CVD

Health surveys estimating the levels of CVD risk factors (and allied health status measures)

Medical records and administrative data reflecting CVD incidence, events and deaths

Model, using the combined data sources, the number of patients who are at high risk (>= 20%) of developing CVD over the next 10 years

By examining EHRs determine the prescribing history among patients at high risk

Estimate the number of patients who might benefit from statins but are not receiving them

Simulate the reduction in heart attacks among these patients if they were prescribed and adhered to a statin
Cost and Benefit Estimation

- Estimate the cost of inviting for screening all patients estimated to be at elevated 10 year risk of a CVD event such as heart attack
- Estimate the cost of prescribing statins to all patients diagnosed as being at high risk
- Estimate the benefit in terms of hospitalisation and other treatment of events prevented

Scenario

The public health team at Megachester were concerned at the high prevalence of CVD in the area and the associated high premature death rate. They knew that people in their area had high levels of risk factors for CVD, and they wanted to explore the potential for reducing some of this risk by increasing access to statins.

Using national data supplemented with local information on the characteristics of patients in their area they ‘shrank’ some risk factor information down to neighbourhood level. For other risk factors they had more direct measures, for example cholesterol and blood pressure from vascular screening records in EHRs. Missing data was modelled using information derived from relevant research.

They set out various options for increasing access to statins and used a Markov model to simulate the potential reductions in CVD incidence with each option. The options included all people with more than 20% 10 year risk, and targeting to various age and community groups. Monte Carlo simulations were used to reflect the uncertainty in the simulated information.
## Targeting Public Health Interventions to Reduce Obesity

**Decision:** How can interventions to reduce the prevalence of overweight and obesity be targeted to achieve the maximum benefit in my population?

**Background**

The rising prevalence of obesity around the world is well-documented but there is a lack of evidence on sustainable public health measures to tackle this complex problem. Given the high burden of early death and suffering that obesity causes, public health services are duty bound to monitor the problem and make reasonable efforts to tackle it. Any public health service might reasonably ask how to target its scarce resources to those most at risk of obesity. For example, given evidence that low socio-economic status is a risk factor for obesity a public health group might consider targeting deprived communities with physical activity and healthy eating promotion. The evidence, however, is inconsistent: most papers show greater excess body weight among more deprived women; but in men some studies show no social effects, some indicate a positive association with deprivation, and some a negative association. So public health professionals will want to know how overweight and obesity is distributed in their populations.

Relevant data sources might include local obesity surveillance programmes, EHRs and health surveys. For example, the Health Surveys for England (HSE) records height, weight, waist circumference and socio-economic data. This is a general population survey that can be used to profile large populations (1-2M) but it has too few observations to represent smaller areas such as towns of say 0.3M. More densely sampled data sources may share some of the same variables as HSE, enabling statistical models to be built that can make “small area estimates”. In addition, primary care services record heights and weights of patients and the postal code of the residence of the patient can be used to infer socio-economic status. Different methods of estimating average body mass index (weight for height, kg/m²) make different assumptions, and it is good practice to explore different potential biases in this way.

**Typical Data Inputs**

- Health surveys with age, sex, ethnicity, household income, education level, height, weight, waist circumference etc.
- Electronic health records with height, a recent weight, age, sex, a geocode from which socio-economic status can be inferred, relevant healthcare information etc.

- Model the prevalence of obesity by population sub-group in small areas
- Simulate the public health impacts, by sub-population, of interventions to promote healthy energy balance
Scenario

Peter, the weight management commissioner for Mezopolis, received the latest figures for his population from the national obesity surveillance scheme. The statistics placed Mezopolis as one of the worst areas in the country for excess body weight. Peter was concerned that this oversimplified the challenges that his relatively deprived community faced. So he asked the public health team to analyse the pattern of BMI by deprivation group.

National surveys had sampled just enough households in Mezopolis to relate BMI to deprivation and sex. Consistent with scientific papers on the subject, women from the lower income households tended to carry more excess weight. Unlike the literature, however, the overweight and obese men in Mezopolis tended to have higher incomes. Peter and the public health team decided that a more detailed picture of BMI was required, including ethnicity and neighbourhood patterns.

The national survey data was insufficient for profiling small groups and areas. The only unbiased data on BMI at this level was from child health surveys in the local schools. Adults are measured by their primary care services, but the fatter adults tend to be measured more than the slimmer ones, creating a biased picture. Some healthcare contacts, such as new patient registration in primary care, antenatal clinics, occupational health screening and vascular screening, lead to routine measurement of weight in some groups. So an adult obesity profile for Mezopolis would have to be estimated using a variety of assumptions and models.

A statistician in the public health team considered using a “synthetic shrinkage” method to produce small area estimates of obesity prevalence from the national survey data linked to some detailed profiling of local neighbourhoods. He thought that the assumptions of this modelling were difficult to justify so he did not proceed. Instead he searched primary care records for contacts where weight was known to be recorded routinely. Three different models for estimating adult BMI by neighbourhood were used, but they produced a similar ranking of neighbourhoods by mean BMI or obesity prevalence.

The neighbourhoods were categorised by BMI, deprivation, ethnicity and other factors known to affect either obesity risk or the uptake of weight management interventions. The evidence base on the effectiveness of interventions to control weight was too patchy to enable a simulation. So a sensitivity analysis using a selection of high quality studies was run in order to explore the value of targeting more weight management resources to specific neighbourhoods. Eventually, a largely qualitative decision was made to target deprived neighbourhoods, mainly because their residents are harder to reach with health promoting activities. It was also decided to improve the surveillance of adult obesity in primary care and some workplaces.
Reconciling Methods of Monitoring Smoking Cessation

Decision: Is my local smoking cessation service as effective as those in other areas?

Background

Tobacco smoking is strongly associated with the CVD and lung cancer. When a smoker quits there is a measurable improvement in their health. When large groups quit, for example after the ban on smoking in public places, there are mass effects including rapid falls in the rates of heart attacks.

Since smoking is highly addictive, the process of quitting smoking is complex, and not simple to measure. There are different sources of information on smoking quit rates, and the characteristics of smokers who quit. This information is either gathered directly from those who use smoking cessation services or it is estimated from population surveys for those quit on their own.

The health risks from smoking are proportional to the amount smoked and the period of exposure. The ideal risk measure is sometimes termed “pack years”. The smoking data in EHRs, however, may be much cruder – for example a code for “ex-smoker” referring to someone who quit last month but was smoking for 30 years, alongside another person with the same code who quit 20 years ago after only smoking for two years.

Both primary care and public health services are paid to help smokers to quit. National health bodies need to monitor the quality of these services and consider how they should evolve as the characteristics of smokers change over time.

Typical Data Inputs

- Estimates of smoking quit rates from national surveys
- Estimates of smoking quit rates from EHRs
- Estimates of quit rates from smoking cessation services
- Model the overall quit rate for the district
- Model the effect on CVD, lung cancer and all-cause mortality rates of changes in smoking quit rates
Scenario

Sarah, who works in Metroville’s public health department, wanted to estimate the local smoking quit rate so she could forecast the impact on CVD and cancer rates. From reading the literature she was familiar with the characteristics of the smokers who quit. She used this information, along with intelligence from patient records and the quit smoking team, to construct a regression model to forecast the annual smoking quit rate in Metroville and its effect on CVD incidence.

Sarah estimated the prevalence of smoking in Metroville by calculating a synthetic estimate from a combination of national and local data. She used patient EHRs and data from the smoking cessation service to estimate the annual number of smoking quitters. The model was revised and updated on a continuing basis as more data were collected.

From the estimates of smoking prevalence and quit rates Sarah constructed a set of forecasts of the effect on the incidence of CVD and lung cancer. Although many of the factors associated with quit rates can be estimated from demographic and socio-economic data, e.g. age and social class, one crucial variable is the strength of the intention to quit. In the models Sarah varied the value of this variable and observed its effect on quit rates. It became clear that crude statistics on quit rates are inadequate for comparing one smoking cessation service with another.
Mapping Missed Opportunities to Prevent Cardiovascular Events

**Decision:** Which general practices should be alerted about the quality of their blood pressure control in respect of CVD care pathways and events?

**Background**

This vignette and the related Appendix I outline a method for identifying sub-optimal treatment of patients who have experienced a cardiovascular (or any other chronic disease) event.

Clinical guidelines or care pathways can be considered ‘idealised’ or ‘expected’ care. Such pathways provide a framework for searching EHRs to reveal the care pathways that are recorded for patients – i.e. ‘observed care’. This provides computable entities for contrasting observed with expected care in order to monitor the quality of services. CVD events such as heart attacks can be considered trigger points for such analyses. In other words, triggering the question “were there any healthcare opportunities missed for preventing this event?” Clinical audits might be expected to ask such questions but they seldom operate across boundaries such as that between primary and secondary care. In reality, the outcomes of care from a cardiology service in a hospital are affected by primary care such as vascular screening and risk reduction. Thus ‘missed opportunity mapping’ needs to consume health data integrated across populations.

**Typical Search and Analysis Workflow**

- **Guidelines** → Identify the risk factor targets which clinical guidelines recommend for optimal care and which are available from the EHR → **EHR**

- **EHR** → Identify from EHRs the patients who had experienced a CVD event

- **EHR** → Identify the patients whose CVD risk factor measurements did not meet targets, and the extent to which the target had been missed

- **Literature** → Estimate (from a model) the number of cardiac events which might have been prevented if the targets had been achieved
Scenario

Dr Jones, the lead commissioner for cardiovascular healthcare services for Megachester, suspected that the unusually high variation in CVD events across the district might be due to differences in vascular screening and risk control.

Dr Jones asked her informatics department to map the guidelines and care pathways for coronary heart disease (CHD) to EHR codes. They reported that blood pressure and cholesterol control in primary care was easily measured for patients with a CHD diagnosis. This information could be linked with hospital admissions for heart attack or acute coronary syndromes, and to deaths data. For all patients experiencing a CHD event their prior blood pressure and cholesterol levels were analysed – 30% had poorly controlled blood pressure and 20% poorly controlled cholesterol. Many patients were also missing these measurements in the year before the event.

The blood pressure and cholesterol control was then analysed for all patients with CHD, irrespective of whether or not they had experienced a cardiac event. Patients from some practices were found to have tighter risk factor control than others. On further analysis it was noted that patients who'd experienced a missed opportunity tended to be from the more economically deprived areas of the district. There was also a higher proportion of missed opportunities among the older patients.

Dr Jones used these findings to persuade her primary care colleagues to review the treatment of all patients with elevated risk factors for CVD, and to ensure that older patients and those living in the more deprived areas received close attention.
Monitoring the Diagnosis and Treatment of Heart Failure

Decision: What changes can be made to local services in order to optimise the treatment of patients with heart failure (HF) – ensuring that all health professionals, the patient and carers have the information they need, particularly over medication?

Background

Heart failure is a complex syndrome of impaired cardiac function caused by structural or functional abnormalities of the heart. The condition occurs predominantly in older people. For around 70% of patients the underlying cause is coronary heart disease. Appendix II describes shortcomings and inadequacies in the diagnosis and treatment of people with HF. The scenario below outlines a method of information collection and management which could improve the co-ordination of the treatment HF patients receive.

Scenario

Dr Ellen Jackson, the CVD lead at the Buena Vista healthcare centre, wanted to improve the information in the EHR available on the diagnosis and treatment of the practice’s patients with or at risk of heart failure. She was aware of the inadequacies of the current information collected, and drew up a specification for a more effective method of ensuring patients received the optimal medication. The following are the requirements that Jackson identified:

**Diagnosis**

Follow the NICE guidelines on the diagnosis of HF, and ensure each stage is recorded. If any patient has their name on the HF register on a provisional basis, remove the name from the register if there is no evidence of HF.

For patients diagnosed in hospital with HF after admission, ensure that the diagnosis is received at the practice and entered on the patient’s record. Specialist services should include the appropriate primary care read codes in order to ensure that patients are coded correctly on systems. Add the patient’s name to the HF register.

Ensure that the information received from the hospital relating to the admission is added to the record using agreed protocols.

**Treatment**

The prescribing of ACE inhibitors and beta blockers for patients with HF is variable, especially the up-titration of medications to target doses. The information system should alert clinicians to those patients who may not be receiving the appropriate medication at an appropriate dose.

**Long-term Condition Reviews**

Patients with HF typically have more than two serious health conditions. These other morbidities affect the treatment which is appropriate for those patients. Such reviews should not be held
separately for each condition. Reviewing all the patient’s morbidities at the same meeting will help to ensure appropriate treatment is offered, and contra-indications taken into account. All the data presented at those meetings should be recorded consistently.

**Typical Workflow for Monitoring Heart Failure Care**

Patient seen by primary care clinician with HF signs and symptoms

Investigation for HF: BNP test and echocardiogram

HF diagnosed: Enter patient details on heart failure register

Record all treatment provided to the patient using clinical terms and codes understood by ALL members of the team providing care from different parts of the healthcare system

Monitor treatment. Assess regularly for signs of decompensation

Patient admitted to hospital with e.g. breathlessness

HF suspected and patient examined

HF diagnosed – send reports to be added to patient’s primary care record. Specialist services should include primary care codes
**Verifying the Accuracy of a Heart Failure Register**

Search the HF register to verify that patients have a confirmed diagnosis of HF and are receiving appropriate treatment

If no evidence of HF found remove name from register. Investigate patients for other conditions

Where evidence for HF is inconclusive, test patient for signs of HF, using BNP and echocardiogram

If HF confirmed ensure patient is receiving appropriate treatment

**Ensuring the Identification of Patients with Heart Failure**

Search EHRs for evidence of patients receiving diagnostic tests or treatment for HF e.g. ACE inhibitors

If sufficient evidence, add those patients to HF register and ensure they receive appropriate treatment

If evidence inconclusive, invite patients to clinic to test for HF

If no signs of HF investigate reasons for medication

If diagnosed with HF, add patients to HF register and ensure receive appropriate treatment
**Patient Co-produced Information for Cardiovascular Health(care)**

**Decision:** How can my practice leverage Web and mobile technologies to improve the management of people with long-term conditions, particularly CVD?

**Background**

Clinicians can prescribe only partial treatment for CVD, full treatment requires patient participation. For example, after a heart attack, moving from the typical diet of Glasgow to that of Thessaloniki is as effective at preventing the second heart attack as taking statins – making both this lifestyle change and taking the drugs affords maximum benefits. Taking regular physical exercise after a heart attack is also vital to preventing a further event. Exercise is also important in preventing adverse consequences in heart failure, which may be counter-intuitive to patients, who are frightened of causing further damage to their heart by stressing it. These vital parts of care, which can only be delivered by the patient in their daily life, are seldom recorded properly in health records.

At the same time as medical records are suffering from patient recorded information, the use of Web and mobile technologies to record aspects of daily life is becoming the norm. For example, commonly available bathroom scales will transmit weight via WiFi and cloud services to mobile phone apps – all the user needs to do is stand on the scale – the device and Web service work out which family member is being weighed. The target market is the ‘worried well’, pushing down the price of this technology well below that of medical devices. For the patient with HF, daily weight measurement is extremely valuable in monitoring the condition in order to make adjustments to drug treatment. Similarly, ‘wellbeing’ applications for monitoring physical activity are relevant to supporting cardiac rehabilitation.

Connections between consumer health & wellbeing products/services and medical records might be facilitated by current policies to give patients on-line access to their records. The target for this access in the UK is universal primary care record access for patients by the end of 2015, and for the EU a 20% access.

**Scenario 1**

Mrs Brown is a 76 year old lady with early HF and a history of coronary heart disease. She saw a poster in her primary care physician’s office about accessing her records from the Web, so she gave consent for this to be switched on. The consent included linking to external services from approved suppliers, including companies making connected bathroom scales and pedometers.

Mrs Brown asked her daughter to help her access her record via the Web. On doing so she noticed that a medication she no longer collects from the pharmacy was on repeat prescription. She also noticed that “none” was recorded under “allergies” when she remembered being very ill after receiving penicillin many years ago. So she used the comments box to make notes about this and her clinician called to discuss and make corrections.

Mrs Brown’s daughter bought her the weighing scales and pedometer for her birthday. She enjoyed using these as she felt more involved in her own care, and had practical targets to aim for. The HF
nurse told Mrs Brown about an application from a medical charity that connects people with HF with one another to compare progress with lifestyle targets such as walking more and to compare experiences with monitoring symptoms such as the weight gain due to fluid retention when drug treatment is out of kilter. Joining in this social network helped Mrs Brown to feel less like a passive patient and more like an active customer of ‘wellbeing in spite of having HF’.

**Scenario 2**

As part of his clinic’s vascular screening programme Dr Schmidt approached a local company about designing a mobile app to enhance the experience of his patients with screening. Initially he wanted to have more accurate CVD risk information available close to the time of screening. As discussions progressed he became more interested in the potential of this technology to engage patients in risk factor control long after their contact with the screening programme. The app company raised the use of social media such as Facebook and Twitter. Dr Schmidt read some papers on cognitive models of health promotion in weight control and realised that effective intervention needed a less medical and more wellbeing approach. So he discussed a bigger project with the CVD service commissioners and public health team in his area. Together they decided to pilot the use of a vascular wellbeing app that would be offered as part of the invitation to vascular screening, and they linked the necessary consent process to the extant scheme to open primary care records to patients. There were some initial difficulties over validating the algorithms for calculating CVD risk but the pilot was very popular among patients. Dr Schmidt eventually took over the primary care aspects of the app to use it to help people with raised CVD risk to manage their condition – including automatic alerts sent to him when patients on CVD care pathways are off target. The public health team and a group of patients eventually took over the ‘pre-clinical’ aspects of the app – maximising its appeal as a fun, wellbeing tool.

**Population Summaries**

In both of the scenarios above social network features were key to engaging individuals in monitoring their health. The social information may be with anonymous reference (to ‘the average patient like me’) as well as supporting peer-to-peer communication. Thus population level searches of the data and modelling are key – and this is an ongoing process to keep apps ‘fresh’ and to quality assure the medical aspects of the information.
**Real-world Trial of a New Medicine to Prevent Stroke**

**Decision:** Should my health economy invest in a new, expensive medicine that provides a more convenient way to stop blood clots forming in patients with abnormal heart rhythms, thereby preventing strokes?

**Background**

People with abnormal heart rhythms, particularly atrial fibrillation (AF), are at increased risk of stroke because blood clots may form in the heart and travel to the brain, blocking small arteries. To reduce this risk patients are usually given the drug warfarin, which generally slows down the clotting process. Getting the dose of warfarin right, however, requires monitoring with blood tests. Too little warfarin and it does not prevent strokes, too much and it can cause strokes from bleeding in the brain.

Some newer drugs such as dabigatran do not require so much monitoring as warfarin does, so they may be safer and more convenient for the patient. But the new drugs cost more than warfarin (including its cost of monitoring). Clinical trials of the new drugs compare them with warfarin, but some patients, particularly those with multiple conditions, have been excluded from these trials. So, the commissioners of CVD services are left with a lack of evidence about how any potential investment in alternative drugs to warfarin will work for their patients, many of whom have multiple conditions.

At the same time, medicines regulatory agencies are calling for more “real world evidence” of the clinical effectiveness and cost effectiveness of new drugs. In future, drugs may be given a partial licence after passing conventional trials, until the picture emerges of how they are working in the complex real world of healthcare, not the artificially simplified world of clinical trials. In order to achieve this EHR data will need to be reused and linked to research data.

**Scenario**

Drug GK5777 has just been given a license and is due to be marketed as “Thinotran”. The business case for this drug relies on adherence being better because it is more convenient for the patient to take – it is taken once per day, at any time, can be taken with most other medicines, and does not require regular blood tests to get its dose right.

Healthcare commissioners/payers are concerned that the cost of the drug will be higher than the conventional warfarin plus the associated monitoring. The current Thinotran trials exclude people with multiple conditions, but searching EHR data reveals that over half of the people with a code for AF and a warfarin prescription have more than one long term condition. So the payers and the drug company decide to run a “real world”, open-label study where patients are randomized to the new drug or warfarin, but both patients and researchers know which drugs are being given.

The local stroke service redesign group become involved in setting up the new Thinotran study. They raise the relation between deprivation, carer support and adherence to medication among the typical elderly person who may be offered Thinotran. The initial study design used postcode linkage
to infer deprivation from a patient’s area of residence. The stroke team then pointed out that deprivation scores from a national database were unreliable for the area where the study was due to take place, so more detailed socio-economic measures were incorporated. The stroke team also raised concerns over some of the assumptions being made by economists modelling the costs of stroke due to failure of warfarin titration.

EHR data were eventually linked with: the clinical trial information system, financial systems that record healthcare transactions/costs, pharmacies, and social care information systems. Patient reported measures of quality of life were also used via mobile devices as well as research nurses – this provided a richer longitudinal picture for the subgroup able to take part. The eventual analysis was more inter-disciplinary than usual, and drew heavily on local healthcare knowledge as well as data available from databases.

In mop-up meetings after the trial, the drug company, payers and the healthcare providers involved agreed that the costs of future studies like this could be reduced by reusing the data linkages and overall information system that had been established for the trial. Furthermore, all parties acknowledged that such an information system would support the quality improvement activities of the healthcare system substantially.
Summary and Conclusions

In this report we have outlined typical requirements for summarising health records at the population level, using examples in cardiovascular healthcare.

We have considered five population level uses of health information for preventive and early healthcare:

1) public health and social interventions;
2) clinical audit and optimising healthcare services;
3) payer evidence and commissioning healthcare services;
4) consumer health applications;
5) research using health records.

We have used a series of vignettes in early and preventive cardiovascular healthcare to illustrate the functions of population health information. Central to each vignette was decision that needed to be taken for more than one patient.

Conventional public health decisions were considered in respect of tackling obesity, promoting physical activity, promoting healthy eating and helping people to quit smoking. Controlling these risk factors was also explored in the clinical contexts of vascular screening and secondary prevention. Parallels were drawn between the lifestyle modifications a person might make to reduce their risk of CVD and the same modifications to prevent a second heart attack. The emerging ubiquity of Web and mobile technologies was presented as a medium to multiply health promotion and secondary prevention efforts, realising substantial health gains for society.

Public health uses of data were also linked to similar population level analyses for clinical audit or healthcare quality improvement purposes. Current health systems usually build separate pipelines of intelligence for: commissioning and financial control; clinical audit and quality improvement; public health; and research. All of these uses are challenging in terms of exploiting linked, and potentially linked, data to their full utility. At the same time, human resources for analysis are scarce and spread increasingly thinly across more data and more requests for information. There is a pressing need for semantically interoperable data transformation processes to enable different users of health information systems to borrow strength from one another.

The scenarios in this report are intended to give health informaticians and software engineers a selection of use-cases for summarising health records across multiple patients. These scenarios highlight the importance of information systems to enable better preventive and early healthcare – a strategic priority for Europe. The appendices provide a link to more technical detail in respect of the informatics. Further work in SemanticHealthNet will expand on the semantic technologies required to assure that population level summaries of health information are consistent across different parts of Europe and different time periods.
Appendix I: Missed Opportunities Mapping

Missed Opportunities Mapping: Computable Healthcare Quality Improvement

Benjamin Brown, Richard Williams, John Ainsworth, Iain Buchan

Centre for Health Informatics, Institute of Population Health, University of Manchester, UK

Abstract

Introduction: Analysing variance from care pathways in situations when adverse health outcomes have occurred may identify missed opportunities for healthcare improvement.

Methods: We developed a computational model for contrasting observed with expected care in pathway searches of coded electronic health records (EHRs). The model was applied in Salford, UK, looking at blood pressure (BP) control and cardiovascular disease (CVD) events. BP was summarised as the integral of serial measurements.

Results: A missed opportunities mapping (MOM) model consisting of a graph of disease events and pathophysiologic states was used to articulate all CVD scenarios conceived. In 3718 patients suffering CVD events in Salford (2007-2012), 1186 (32%) had suboptimal BP control. This missed opportunity detection rose to 36% using the integral instead of the most recent BP record.

Conclusions: MOM provides a useful, computable model for encoding care pathways and searching EHRs to detect variations from expected care. Further research is needed in other disease areas. The indications however, are that this model could be used to embed healthcare quality improvement at both patient and population levels.

Keywords: Integrated Care Pathways; Guideline Adherence; Medical Records Systems, Computerized; Quality Assurance, Quality Improvement.

Introduction

Background

Given the ubiquity of clinical guidelines and the increasing prevalence of coded electronic health records, there are potentially computable entities of ‘ideal’ care for many healthcare situations [1]. However, there is often discrepancy between ‘ideal’ and what happens in real-life clinical practice [2].

When adverse outcomes have occurred (e.g. myocardial infarction; hospitalisation; death), identifying instances where real-life clinical care has avoidably deviated from its ideal are potential ‘missed opportunities’. Further analysis may provide vital information on how healthcare services could be improved to prevent future adverse events. If this information can be provided in a cost-effective, timely and automated way it might support better targeted decision-making by policymakers, providers and payors. Furthermore, patients and clinicians will have the foundation of a system for highlighting and helping to avoid missed opportunities before they lead to adverse outcomes.

Aim of the paper

In this paper we present MOM – a new theoretical model of how to identify missed opportunities in healthcare to support quality improvement. We present a computational model for MOM and report its application to CVD and BP analysis with coded electronic health record data from the UK.
Importance and relevance of the paper

MOM is a new approach to healthcare quality improvement using an informatics approach that has not yet been reported. The technique employs two unique methodological features which differ from the existing literature with respect to analysis of guidelines at the population level, and how to represent pathophysiological measures.

This paper also has direct clinical relevance. By preventing adverse outcomes through more effective upstream management, MOM provides a new way to contain healthcare costs amid the increasing burden of chronic diseases. In addition, by enabling in-depth analysis of different patient groups suffering missed opportunities, MOM can be used to address the problem of widening inequalities in healthcare.

Materials and methods

First we present the theory behind MOM, followed by the computational model and software involved, then its application to integrated health records from Salford, UK.

Missed opportunities as a concept

Ideal care is encoded in clinical guidelines as rules that define appropriate or high quality care [3]. Care pathways contextualise guidelines by translating their abstract statements into idealised patient journeys [4]. At the other end of the spectrum, the most accurate representation of real-life clinical practice is captured in health records. Understandably though, fidelity is lost depending on a variety of user and system-related factors. Missed opportunities therefore lie in the theoretical gap between clinical guidelines/care pathways and healthcare records (Figure 1).

By analysing the variance between clinical guidelines/care pathways and healthcare records, we can thus quantify and characterise missed opportunities. This provides information regarding instances along patient journeys where accepted best practice did not occur and may have contributed to an adverse health outcome (Box 1).

Figure 1 – Where missed opportunities arise

Box 1 – Generic definition of a missed opportunity

Computational model and software

We used software already developed by our group to handle the data (COCPIT; Collaborative Online Care Pathway Investigation Tool) [5]. COCPIT is a flexible Web-based tool for exploring Electronic Health Records (EHR). The front-end is a Silverlight plug-in accessed via common Web browsers, and the back-end is a combination of an SQL Server database and C# .NET web service. The basic function of the tool is to define Events and States. An Event is a single point in time occurrence such as a diagnosis or clinical measurement. States are conditions that define a group of patients based on Events, such as ‘patients who smoke > 20 cigarettes per day’ or ‘patients with diabetes and chronic kidney disease who have had less than 2 blood pressure (BP) tests in the last year’. Events and States are built as queries using terms from the EHR’s coding set. These may then be inverted or combined with Boolean operators to create multifaceted Events or States (Figure 2). COCPIT is flexible enough to deal with any set, though EHRs typically use Read Code v2, CTV3, ICD10 or Snomed [6].
The number of patients in a State can be categorised by various patient characteristics such as sex, ethnicity, age, deprivation (imputed from area of residence and the Index of Multiple Deprivation [IMD] [7]) and number of health conditions (morbidity). Statistical analysis can then be performed.

To enable COCPIT to identify missed opportunities, a State must be defined as the ‘failure to deliver a quality standard of care’ in a clinical guideline/care pathway with the subsequent ‘adverse outcome’ of interest as an Event (Box 1). COCPIT may then capture how many patients are in the state of ‘missed opportunity’ at the point they experienced the Event.

**Demonstration study**

We demonstrate the MOM technique on CVD events (myocardial infarction [MI], stroke and transient ischaemic attack [TIA]). CVD was chosen as it creates massive global economic burden yet is largely avoidable through adequate risk factor management [8]. We focus on missed opportunities in BP control in this paper as it is considered the most important clinical aspect for reducing CVD mortality [9,10]. The MOM model has been tested in a wider range of contexts but space limits reporting them here—indeed MOM is designed as a generic methodology.

The study was set in Salford, UK, which has a population of 0.22M and is relatively deprived [11]. Data were extracted as Read Code v2 terms from the Salford Integrated Record (SIR) – the UK’s first fully integrated EHR between primary and secondary care. Patients who experienced their first MI, Stroke or TIA between April 2007 and 2012 were eligible. Data were cleaned by removing non-numeric values and those with dates after 2012.

To determine the missed opportunity criteria for the demonstration study, we consulted an expert clinical panel. They recommended guidelines advising a BP of <140/90 mmHg to prevent CVD events [12–14], though also acknowledged other recommendations such as the UK’s Quality and Outcomes Framework (QOF) target of ≤150/90 mmHg [15]. They also cautioned that instances where it was clinically inappropriate to lower BP, such as in those with recurrent falls, should not be counted as missed opportunities. A typical missed opportunity is shown in Box 2.

**Box 2 – Definition of a typical missed opportunity**

Statistical results are presented as main effects and 95% confidence intervals where practical. An exact binomial method is used for confidence intervals of proportions [16]. Liddell’s test is used to compare pairs of proportions arising from using different methods to analyse the same patient records [17].

**Results**

**Computational model and software**

Using COCPIT we built a query that described patients with ‘uncontrolled BP’ (Box 2) as a State and subsequent MI, stroke or TIA as an Event. The Read Codes used were: myocardial infarctions (G30%); stroke (G61%; G64%; G66%); TIA (G65%); systolic (2469) and diastolic (246A) blood pressure; chronic kidney disease (CKD; 1Z1%); and diabetes mellitus (DM; C10F). To account for ‘reasonable clinical exceptions’ (Box 2), we excluded patients where lowering BP may not be possible due to factors such as palliative care, recurrent falls, orthostatic hypotension and maximal tolerated therapy. The Read codes used were: under the care of community palliative care team (9Nh0.); recurrent falls (16D1.); orthostatic hypotension (G870.); and patient...
on maximal tolerated antihypertensive therapy (8BL0). In addition we used multiple codes for different antihypertensive medications and their refusal or contraindication.

To determine whether a patient’s blood pressure (BP) was uncontrolled we developed an alternative to just considering the most recent measurement, which is standard practice for most clinical guidelines, targets and research [18,19]. Considering that the inherent variation in BP, due to a range of external factors [20], could lead to a single reading being over interpreted, we decided to incorporate information from series of BP measurements. In order more accurately to reflect a patient’s BP over time we calculated the time-dependent average, which is the integral of BP readings over time divided by the time period in question as shown in Equation (1).

\[
\text{Time-dependent BP average} = \frac{\int_{\text{Date of MI, stroke or TIA}}^{\text{Date of first BP reading}} f(BP) \, \text{/ Time}}{1}
\]

This provides a value with the same units and range as a usual one-off BP reading, with the added information of multiple readings and their variance over time. The integral of the BP readings equates to considering that a patient’s BP has changed linearly between measurements or over rapid changes at the midpoint of measurements (Figure 3).

**Demonstration study**

Between April 2007 and 2012, 3718 Salford patients suffered MI, Stroke or TIA. Over half (2163, 58%) were male. Most ethnicity data was unrecorded (2401, 65%), though the vast majority that were recorded were white (1263, 34%) rather than non-white (54, 1%). The median deprivation score was 34 (IQR 22-50), and the median age at adverse event was 71 (IQR 59-80) years. The last BP measurement was taken a median of 79 (IQR 21-213) days prior to the adverse event.

According to our preferred quality standard of BP <140/90 mmHg, 1186 (32%) out of the 3718 patients experienced a missed opportunity when considering their most recent BP measurement. This increased to 1323 (36%) when the time-dependent average BP was used (P<0.0001). This figure compares to 647 (17%) according to the QOF target of ≤150/90 mmHg (P<0.0001).

Figure 4 details the proportion of patients who suffered a missed opportunity according to sex, ethnicity, deprivation, age and morbidity (CKD, DM or CVD). Stated proportions are in relation to the total number of patients in that subgroup. The null hypothesis was that each subgroup proportion suffering a missed opportunity would be the same as the overall population proportion suffering a missed opportunity. For example, if one-third of patients suffered a missed opportunity, then it would be expected that one-third of those aged 35-44 years (or any other characteristic) would also suffer a missed opportunity.

The main points from Figure 4 are that firstly, there is no statistically significant difference in proportions of patients suffering a missed opportunity according to gender or deprivation. This implies that neither characteristic is associated with missing an opportunity. Secondly, patients of white ethnicity suffer significantly less missed opportunities, though this should be interpreted with caution due to the large amounts of missing data. Thirdly, there is a clear statistically significant increase in the proportion of patients suffering a missed opportunity with advancing age. This suggests the older a patient the more likely they are to suffer a missed opportunity. Finally, patients with none
or more than one clinical condition are significantly more likely to suffer missed opportunities than those with just one condition. Each of these associations may be explained by a variety of patient, clinician or system-related factors.

Figure 4 – Proportions (with 95% confidence intervals) of missed opportunities in different patient groups

These findings suggest immediately applicable clinical recommendations to reduce the burden of CVD in Salford. Firstly, they highlight that a large proportion of patients are suffering missed opportunities in their clinical management prior to suffering CVD events. Secondly, they raise the question of which BP quality standard should be aimed for given the statistically significant difference in proportion of patients suffering a missed opportunity according the different targets (<140/90 vs. ≤150/90 mmHg). Finally, they suggest that particular patients are at higher risk of suffering missed opportunities – namely elderly patients and those with none or greater than one clinical condition. Arguably these patients should therefore be targeted for more intensive BP reduction in future, which may be facilitated by addressing the causative patient, clinician or system-related factors previously alluded to.

Discussion

This paper has presented a theoretical and computational model of how to identify missed opportunities in clinical care from coded EHR data. The demonstration study has also shown how this model could work using real-life data from the UK. This system could be applied to any routinely collected primary care data, not just the early adopters of integrated primary and secondary care records. We have demonstrated that missed opportunities are common and that ‘mapping’ them can provide important information to improve clinical services.

There are many more potential uses of MOM. For example the number of potentially avoidable adverse events could be quantified by theoretically eliminating the missed opportunities. This could be done by estimating the relative risk reduction associated with ideal clinical management, such as reducing BP [21]. More accurately this could be achieved by using MOM to inform a retrospective cohort study, comparing the rates of adverse events between those who have and have not suffered missed opportunities. MOM data could also be refined to provide information on the performance of smaller clinical units such as hospitals, GP practices or even individual clinicians. This could target healthcare quality improvement resources in a more agile way than at present.

Using BP and CVD in the demonstration study points to the variety of clinical situations to which MOM could be applied. Any number of CVD risk factors could be substituted for BP, including cholesterol levels, diabetic glycaemic control or smoking. Additionally, other potentially avoidable adverse outcomes could be used such as exacerbations of chronic conditions, healthcare-associated infections, or more general events including death and hospitalisation. Other patient characteristics could also be studied depending on the richness of the data available in the EHR. Furthermore, our comparison between the two BP targets (<140/90 or ≤150/90 mmHg) shows that different quality standards can be compared to consider the ‘best fit’ to local circumstances.

There are a number of distinguishing features in this paper with respect to the existing literature. Firstly, previous health informatics studies of clinical guidelines/care pathways have tended to focus on the prospective care of individual patients. MOM examines
retrospective adherence to guidelines/pathways at the population-level. This is arguably more accurate as guidelines are designed for average patients, calculated from studies with large cohorts, not individual patients [22]. Furthermore, this approach provides more useful information on the performance of a health system and long-term outcomes rather than guidance on individual patients. Secondly, we have introduced a new method for calculating long-term BP measurement, which we consider more accurate and sensitive than current standard of using the latest reading. In our demonstration, the time-dependent BP average identified 137 (4%) more missed opportunities than using the last BP reading alone. These are instances that would have been missed otherwise and were detected by this method because it accounts for patients who have only recently had their BP controlled.

Furthermore, in our study the last BP measure was taken a median of 79 days before the event with a wide inter-quartile range of 21-213 days. It is remarkable that using a BP reading taken so long before the event could provide reliable information on the overall state of BP control. This technique of representing long-term pathophysiology can be translated to analogous clinical scenarios such as monitoring cholesterol, glycated haemoglobin, and estimated glomerular filtration rate.

As with all studies based on routinely collected data, a weakness of this paper is its reliance on the quality of information within the EHR. There is also a lack of qualitative insight that may have revealed acceptable clinical reasons for missed opportunities. However, by taking into account the clinically relevant exceptions (such as recurrent falls) advised by our expert clinical panel, we believe these have been mitigated against.

Ultimately, MOM currently only provides theoretical advice. Therefore to further understand the utility of the concept we need to test its validity in real-life clinical evaluations. This could be done in a pragmatic randomised control trial in the future.

**Conclusion**

This paper presents a new computable approach to healthcare quality improvement, which could be used to inform clinical audit, service planning, and wider health policy. We have demonstrated its feasibility in detecting ways to improve CVD prevention using real-life clinical data from the UK. Future work will look at further clinical scenarios and the utility of the system when embedded in real-life healthcare quality improvement activities.

**Acknowledgements**

We thank our expert panel of clinicians: Washik Parkar, Steve Little, Simon Capewell and Tim Frank. We also thank Artur Akbarov for statistical advice.

Williams, Ainsworth and Buchan were funded by the UK National Institute for Health Research (NIHR) as part of the Greater Manchester Collaboration for Leadership in Applied Health Research and Care (CLAHRC).

**References**


[8] O’Flaherty M, Buchan I, Capewell S. Contributions of treatment and lifestyle to declining CVD mortality: Why have CVD mortality rates declined so much since the 1960s? Heart. 2012 [e-pub ahead of print]


Address for Correspondence

Benjamin Brown
Centre for Health Informatics
Institute of Population Health
Room 1.309
Jean McFarlane Building
The University of Manchester
Oxford Road
Manchester
M13 9PL
UK

Distribution

Appendix II: Heart Failure in Primary Care

Are We Delivering Evidence-Based Care for Heart Failure in Primary Care?

Authors: Christi Deaton, Lorraine Burey, Michael Spence, Joanne Thomas, Kieley Wild, Washik Parkar, Ivan Benett

Aim: Many patients with heart failure (HF) in the UK are managed in primary care (PC) with referral to specialist services only as necessary. The aim of this project was to compare patient characteristics and the quality of evidence-based care for HF in PC with patients and care provided by co-managed specialist services and PC.

Methods: 13 PC practices were assessed using the Greater Manchester HF Investigation Tool (GM-HFIT) of HF register verification, case finding and skills audit. Patients were considered co-managed by specialist services if seen within the past 12 months. A HF specialist nurse and knowledge transfer associate conducted the assessment. Anonymous data were entered into a SPSS database for analysis.

Results: 469 patients were on the HF registers: 60% were appropriate, 16% inappropriate, and 24% needed further evaluation for diagnosis. Patients with definite HF had a mean age of 73 + 14 years, 42% were female, 64% had hypertension, 37% had atrial fibrillation, and 31% had diabetes. Co-managed patients (28%) were younger, more likely to be male, to have an ejection fraction < 35%, more likely to have certain assessments and education, and if appropriate to be up-titrated to target doses of ACE inhibitors (ACEI) and beta blockers (BB) (Table).

Conclusions: Older patients, women, and those with EF > 35% were less likely to be co-managed. Although a high proportion of PC patients were on ACEI and BB, up-titrations were low. Co-managed patients were more likely to receive evidence-based care, indicating a need to improve management in PC and collaboration between PC and specialist services.

Table

<table>
<thead>
<tr>
<th>Variable</th>
<th>All patients (n=390)</th>
<th>PC only (n=281)</th>
<th>Co-managed (n=109)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean age (sd)</td>
<td>73 (14)</td>
<td>74.5 (14)</td>
<td>70 (14)</td>
<td>.007</td>
</tr>
<tr>
<td>Female</td>
<td>42%</td>
<td>46%</td>
<td>30%</td>
<td>.004</td>
</tr>
<tr>
<td>EF &lt; 35%</td>
<td>49.5%</td>
<td>41%</td>
<td>63%</td>
<td>.004</td>
</tr>
<tr>
<td>ACEI/target dose</td>
<td>89%/59%</td>
<td>88%/50%</td>
<td>93%/74%</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>BB/ target dose</td>
<td>77%/43%</td>
<td>71%/27%</td>
<td>87.5%/69%</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Self-care educ</td>
<td>16%</td>
<td>4%</td>
<td>43%</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Nutritional info</td>
<td>16%</td>
<td>5%</td>
<td>45%</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>
Background

Heart failure is a complex syndrome of impaired cardiac function caused by structural or functional abnormalities of the heart (NICE Heart Failure Guidelines 2010). 30-40% of patients die within a year of diagnosis, but survival and morbidity are improved through evidence-based treatment. The condition becomes more prevalent after the age of 60, but can be hard to diagnose.

The identification and treatment of heart failure (HF) can be grouped into five stages. At each stage the recording and availability of reliable data improves the outcomes and lowers risk for the patient.

Stage 1 – Prevention

Heart failure is a condition which occurs predominantly in older people. For around 70% of patients the underlying cause is coronary heart disease. Since for most patients HF is a condition which develops over time it is important to identify in patients the risk factors for HF, and provide treatment to control these. In addition to coronary heart disease, major risk factors are hypertension and diabetes. Accurate medical histories and awareness of risk for HF are essential.

Stage 2 – Diagnosis and treatment in primary care

Since many of the signs and symptoms of HF are also indicative of other conditions, HF is hard to diagnose. Early diagnosis of HF requires a combination of expertise of a primary care physician, appropriate referral to a specialist and easy availability of certain diagnostic tests (eg X-ray, electrocardiography, echocardiography and B-type natriuretic peptide (BNP) blood tests). The BNP test is not available to all GP practices, but is useful to rule out heart failure as well as to identify patients needing urgent specialist evaluation. Patients with previous myocardial infarction (MI) and signs and symptoms of heart failure are high risk, and need urgent echocardiogram and specialist evaluation (within 2 weeks) (NICE, 2010). Initial treatment for HF may begin with specialist consultation, and systems should support communication between primary care and specialist services. An ideal information system could be the interface between both primary care and specialist services, so that information is shared about patient diagnosis and treatment. Specialists could answer questions from the GP, which could ensure that referrals are appropriate and timely, and that recommended treatments are carried through.

Stage 3 – Admitted patient care

Patients may be admitted to hospital for treatment of an acute exacerbation of the condition. A report published in 2007 found that 80% of new cases were found through hospitalisation and could have been identified earlier (Healthcare Commission, 2007). Patients hospitalised for either de-compensation of their HF, or for other co-morbid conditions, need specialist input into management and discharge planning, and early follow-up after discharge. Lack of communication between hospital and community or outpatient specialists can lead to inappropriate care and inadequate follow-up. These factors contribute to high readmission rates for patients (Phillips et al, 2004 and Yu et al, 2006). Providers should be notified when patients with heart failure are admitted and discharged to hospital regardless of reason for admission. Hospital data could also be accessible from the GP and outpatient services.
Stage 4 – Long-term care

As patients become more frail their need for support increases. All treatments should be recorded reliably. The patients themselves need to be aware of their condition and its prognosis. Information systems could generate individual information for patients in lay language, and should link to systems of tele-monitoring and tele-health if patients need that type of support. HF specialists and primary care should agree indications for a patient being referred back to specialist services. Information systems could include information about worsening HF and other treatment options like devices and when these may be appropriate.

Stage 5 – Palliative care

This includes managing pain and discomfort at the end of life. Because acute exacerbations occur in HF and can often improve, the trajectory for patients nearing the end of life can be difficult to identify. Information systems should include appropriate assessments to help providers, patients and families identify the point at which care should be palliative. These systems could also prompt questions and discussions regarding preferences for end of life care, which should begin earlier. For example, patients with implantable cardioverter defibrillators (ICDs) need discussions with clinicians about if and when the device should be turned off.

NICE has issued a diagrammatic summary of the process of diagnosing heart failure and a diagram of the pathway for treating heart failure (NICE, 2011).

Comorbidities complicate treatment. Reviews of patients with long term conditions should be designed to ensure all conditions are reviewed together so that optimal treatment can be agreed. It is vital to ensure that treatment to improve one condition does not exacerbate another.

There is a low reliability of the information on HF registers held in primary care. There is evidence that a high proportion of patients on registers do not have HF. They have probably been miscoded and could be missing out on treatment they should be receiving. There are patients with HF whose names are not on an HF register. They are also likely not to be receiving appropriate treatment.

There is evidence of inconsistent use of terms and codes across primary and secondary care. When a GP refers a patient for an echocardiogram the results are not consistently added to the primary care record.
The treatment of HF in primary care can show several instances of missed opportunities:

- Patients with HF not on the HF register (so potentially less likely to receive HF reviews and evidence-based care)
- Patients without HF inappropriately on the registers (so potentially missing care for another condition)
- Patients needing further evaluation to definitively establish a diagnosis – so may not have started appropriate therapy.
- Patients not up-titrated to target doses of recommended drugs (so not receiving maximum benefit from therapy)
- Patients not receiving a thorough assessment at clinic visits for signs of early clinical decompensation (increased ankle oedema, weight gain, pulse rhythm check [can identify atrial fibrillation and need for anticoagulation]). Recognising early signs of decompensation can prevent hospitalisation. There is evidence that these are infrequently documented in primary care.

References


Distribution

Appendix III: COCPIT

COCPIT: A Tool for Integrated Care Pathway Variance Analysis

John AINSWORTH and Iain BUCHAN
Centre for Health Informatics, University of Manchester, UK

Abstract. Electronic Health Record (EHR) data has the potential to track patients’ journeys through healthcare systems. Many of those journeys are supposed to follow Integrated Care Pathways (ICPs) built on evidence based guidelines. An ICP for a particular condition sets out “what should happen”, whereas the EHR records “what did happen”. Variance analysis is the process by which the difference between expected and actual care is identified. By performing variance analysis over multiple patients, patterns of deviation from idealised care are revealed. The use of ICP variance analysis, however, is not as widespread as it could be in healthcare quality improvement processes – we argue that this is due to the difficulty of combining the required specialist knowledge and skills from different disciplines. COCPIT (Collaborative Online Care Pathway Investigation Tool) was developed to overcome this difficulty and provides clinicians and health service managers with a web-based tool for Care Pathway Variance Analysis.

Introduction

Integrated Care Pathways (ICP) [1] are multi-disciplinary care plans that detail the diagnostic and treatment steps for patients presenting with a particular condition. They are used to aid clinical decision-making, as they effectively implement (national) clinical guidelines, with customisation possible at a local level to reflect service provision. They also help to ensure quality standards are met and to reduce variation in practice. In the UK Map of Medicine [2] and NICE Pathways [3] provide ICPs for many different conditions for use within the National Health Service (NHS). ICPs define a chronological sequence of steps, mostly commonly diagnostic or treatment, to be followed in providing care for a patient. An ICP may branch as a result of a diagnostic result and they may be cyclic where repetition of a sequence of steps is required for the maintenance of a health state. There will typically be many paths through an ICP, and each patient will follow one path. Flow charts are the predominant formalism for representing ICPs.

Where patients are cared for by a healthcare service that implements an ICP, the patient’s medical record will capture the details of the patient’s progression along or around the ICP. By comparing the patient’s medical record against the ICP, we can identify where the care given has diverged from the care expected. This is termed Care Pathway Variance Analysis (CPVA). CPVA can be calculated for individual patients and then aggregated according to the group of interest, such as hospital or clinician to provide a view of performance over a defined time frame. This may be further analysed by examining specific patient groups defined by, for example, gender, age, ethnicity or socio-economic status. The core method of CPVA can then be extended to a range of applications: clinical process audit; clinical outcome audit; inequality audit; identifying inefficiencies in provision of care across multiple morbidities; service redesign and economic evaluation of care pathways. Performing CPVA requires (i) the ICP to be defined in a computable form in terms of the clinical codes used in the corresponding medical records; (ii) electronic medical records accessible and available for use in a
way that maintains patient privacy and confidentiality; (iii) computer codes for the analysis of CPVA and presentation of the results. For an organisation to perform CPVA considerable investment in time and people with specialist skills is required. Consequently, despite the considerable benefits that can be derived from CPVA, it has not been widely developed [4], [5].

In this paper we describe the functionality of a tool named COCPIT that begins to address the issues with CPVA identified above and report on some first experiences of using the tool. Our aim was to remove the barriers to performing CPVA, by developing an easy to use, web-based tool targeted at an end user community of clinicians and health service managers.

1. Methods

Throughout we used an agile, user-centred design methodology beginning with initial requirements capture using paper-based storyboards progressing through successive iterations of software development and review. A user reference group was established with representatives from primary care, secondary care, public health, service managers and commissioners to ensure a balance of views and requirements, and this group provided feedback on the user interface and prioritisation of tasks for the next iteration. COCPIT has three principle components: first, a data management framework, providing access to individual-level medical records whilst ensuring appropriate information governance; second, a visual editor for designing ICPs, which can then be stored; and third, an analysis and visualization component implementing the methods and techniques for users to perform CPVA and summarise the results for a range of applications.

1.1. Data Preparation

COCPIT was developed as a plugin tool for the eLab framework [6]. The eLab is a web-based software framework for data-intensive health research and knowledge generation. It provides secure access to health datasets and enforces information governance policies. Through the Research Object (RO) mechanism it provides the capability for users to record all the activity of their investigation such that it can be repeated or reused by others. For example, a RO can encapsulate an ICP definition that can then be discovered, copied, modified and employed by other users. The eLab content discovery mechanism employs metadata on ROs, combined with social networking techniques to actively suggest content for users. The eLab also provides data pre-processing tools that can reformat patient level data into a format that COCPIT analysis subsystem can understand. For example, transforming a journal table of clinical events for multiple patients to a table containing a list of clinical events for each patient. COCPIT imposes minimal requirements on the input data, but each event should be time stamped so that it can be chronologically ordered for each patient. COCPIT is developed in the Microsoft Silverlight web application framework to provide the interactivity required, whilst remaining web-based.

1.2. ICP Representation

In order for ICPs to be computed and persisted, a suitable machine-readable representation was required. Having reviewed the plethora of guideline languages available [7], we concluded that there was no agreed standard and so we should define our own representation to be used internally to COCPIT, since the most important criteria at this stage was to be able to control and experiment with the definition language, such that it was tuned to our exact needs. Eventually, COCPIT could be
extended to read and write other formats to allow reuse and exchange between CPOCPIT and other systems. We chose to represent ICPs as directed graphs, where each node represents a step on the pathway and the edges represent allowed transitions. Any valid path through the graph corresponds to a possible patient journey. Associated with each node is a statement that defines the event(s) that, if recorded in the patient’s medical record, mean that the node on the pathway has been entered. Events may be defined as clinical codes from a terminology (e.g. SNOMED-CT [8]) or may take on any user-defined value. This allows the ICP to be customised to the dataset over which it will operate. Complex combinations of events can be created using logical operators AND, OR, NOT and nested statements are permitted. Complex combinations are needed where there is variation in coding practice or where a terminology set is ambiguous. Expressions may include temporal relations between events, such that a particular event recurs a defined number of times in a defined period. Nodes may have validity conditions, which can determine whether a transition between a pair of nodes was expected or not. These are defined in terms of a conditional expression on an event(s) with an associated value, such that the value is equal, less than, greater than or falls within a range. A node may also have values associated with each exit that defines the proportion of patients that will follow the edge. Edges between nodes may have an associated transition time value that can define either the upper limit on the transition time between nodes or defines the expected transition time. Our initial implementation of the ICP representation used XML Schema, though this was subsequently changed to RDF as XML proved insufficiently flexible.

1.3. ICP Editor

The COCPIT ICP Editor presents the user with a visual canvas on which nodes can be added. For each node, a statement must be defined, and COCPIT provides a visual statement builder that simplifies the construction of complex statements. The statement builder utilises the eLab terminology service to lookup clinical codes from a search string that can be used to populate the event definitions. Edges are simply drawn between nodes on the canvas to build up the graph of the ICP. Transition time values can be optionally added to the edge definition. If a new exit edge is added to a node that already has one or more exit edges, then the exit condition editor is opened that allows the user to define the expression for each edge.

1.4. CPVA Processing

Having defined an ICP and selected a data source the user can now perform CPVA. The quality of real-world data is highly variable and this has consequences for data analysis. There will be a set of patients whose medical record does not completely match a path through an ICP. There could be many reasons for this, for example: (i) the event has not been recorded in the patient’s record; (ii) care professionals did not carry out one or more steps on the ICP; (iii) the patient refused treatment; (iv) the patient did not attend clinic; (v) the event was miscoded; or (vi) the ordering of events was reversed when recorded, which is particularly common when events occur in close proximity in time. COCPIT cannot determine why data is missing, but its matching algorithms must be able to cope with incomplete data to produce the best possible fit between observed and expected. Another implication is that patients may enter or exit a node in the ICP graph without following a defined edge when there is no record of entering a preceding or a succeeding node.
For a set of patient records and an ICP, COCPIT can currently perform the following analysis:

- Count of matches to each node
- Count of matches of unique patients to each node
- The point in time a patient matches to each node

From these values we can also then calculate values such as time between events, and summary statistics for number of node visits and duration.

The results can be presented in different ways:

1. Overlay the results on the ICP graph, by incrementing the count to for each node visited and each transition made. Display the value for each count on the node/edge as a proportion of the total population
2. Generate histograms and Kaplan-Meier plots for the traversal times across each edge.
3. Full node statistics for patients entering, leaving, appearing, disappearing, proportion following branch vs. expected, and transition times.

COCPIT provides the capability to segment the result by population characteristics. Any combination of age (including user defined ranges), gender, ethnicity and deprivation of their local area can be selected to create a data series, so that comparisons between patient groups can be made. This can be used to investigate inequity in the provision of healthcare to sub-populations.

2. Results

The COCPIT tool was implemented over seven development iterations, with input from the user reference group at each stage. This has ensured that the software meets the needs of the end users. The ICP Editor has been used to implement ICPs for the Greater Manchester Stroke Service and for Chronic Kidney Disease (CKD) in the English NHS. The system has been tested with simulated patient data and real world patient data from the Salford Integrated Record [9] database for patients with CKD. The tool has been successfully run over a database of 100,000 unique patients with a total of 2.5 million events,. COCPIT was applied to the survival analysis of stroke patients at Salford Royal Foundation Trust, UK. The aim was to verify that COCPIT would produce the same results as produce by a statistician using the R statistical analysis software package independently. As the input data set did not use a clinical coding system, the meta-data describing allowed values was extracted prior to loading into the eLab. There was one difference in the results, caused by an inconsistency in the way that R and Microsoft .net handled daylight savings time changes.

3. Discussion

A limitation of COCPIT is that it depends on high quality data to produce meaningful results. Further exploration of matching algorithms is required to improve the handling of missing data. However, this is a potential application for COCPIT that was not previously envisaged: improving data quality and clinical coding of electronic health records. It is ideally suited to this task, as an ICP definition essentially defines what should be coded and how. Future development will provide the functionality in COCPIT to calculate the financial costs of variance from expected care, evidence that can be used to support the case for service redesign. COCPIT will also be extended to show how a service changes over time, supporting a “natural experiment” approach to continuous quality improvement,
especially where multiple health systems compare and contrast their evolution using this tool. We envisage that COCPIT could be used to explore care quality at multiple levels: patient; clinician; clinical team; provider; health system/community; and profession. Eventual machine-to-machine communication of such quality metrics will have profound implications for managing health systems – the ethical and legal issues need to be examined.

4. Acknowledgements

Ainsworth and Buchan are partly funded by the NIHR CLAHRC for Greater Manchester.

References

8 http://www.ihtsdo.org/ (Accessed 28th January 2012)

Distribution

Paper read at MIE 2012.

Appendix IV: Sharable Simulations of Health Policies

Sharable simulations of public health for evidence based policy making

P.A. Couch, J. Ainsworth, I. Buchan
Centre for Health Informatics, University of Manchester, UK.
pacouch@manchester.ac.uk

Abstract

There is a lack of evidence based health policy making due to the difficulty of simulating the potential impacts of policies in locally relevant ways. Here we address the inaccessibility of high quality models of public health and policy - presenting the concepts of a new simulation framework, IMPACT, built on Semantic Web principles. Model and simulation data are persisted with rich semantics and context to support sharing and interpretation. For this purpose, graph storage systems are explored alongside a new framework for mapping clinical data objects to graphical models. The computation employs functional programming for the parallelised simulation of locally representative populations/cohorts changing over time. The input data, model information and simulation results are mapped to social networks of policy making using the Work/Research Object and e-Lab paradigm that is emerging in E-Science. Practical applications to coronary heart disease are discussed.

Introduction

It is recognized that CVD is one of the leading economic burdens to the UK healthcare system [1]. Changes to the pathways of care used to manage the disease and to preventative measures could impact significantly on the health of the UK population and the cost of the healthcare system. Despite its importance, it remains difficult for policy makers to base healthcare decisions on the latest medical evidence. This is due to the difficulty of simulating the potential impacts of policies in locally relevant ways. In order to achieve local relevance a simulation must incorporate local data and be steerable by local policy makers. In spite of the data being available, policy options are rarely informed by simulations specific to local circumstances. This is due to models being inaccessible or inadequate. Dependence has to be placed on simple models that are not accurate representations of reality and that produce unreliable estimates [2].

The IMPACT system is being developed as part of the Greater Manchester CLAHRC programme, a collaboration between the University of Manchester and NHS Trusts across Greater Manchester. The objectives are to implement a system for creating, executing and analysing the results of simulated public health and healthcare interventions. The system is designed to meet a number of requirements outlined in a survey of attitudes of policy makers to simulations and models conducted by Robinson-Taylor et al [3]. These include versatility, flexibility, transparency, accessibility and collaboration. The system allows users to construct graphical structures that represent the disease pathways of interest. Graph nodes represent disease states and arcs represent the possible transitions between those states. Model developers associate transitions with risks that affect the rate that people move between the disease states. Healthcare interventions can be defined and characterised by their effect on these transition risks. In this manor, users are able to examine what-
if scenarios concerning healthcare interventions. The system has been validated using a new model for coronary heart disease that it shown to accurately reproduce observed mortality rates. Open source software is used and service oriented architecture is employed to ensure a clean separation between components. IMPACT is made available through a web based interface\(^3\) to ensure it is accessible to a wide range of users, including those that have limited administration rights on their computer. The current version of the system is described in detail elsewhere [4].

There are numerous modelling methods that can be employed to simulate policy decisions. The methods adopted need to realistically represent the underlying disease and care processes but also be suitable for visual representation and computation. Model complexity has to be balanced against the ease with which it can be constructed and understood. It also has to be balanced against the computational cost of executing simulations. The UK e-Science Core Programme lead to significant advances in distributed computing paradigms and frameworks, allowing straightforward execution of high throughput tasks. The same is true for the increasing availability of cloud computing resources, such as the Windows Azure Platform and Amazon EC2. It would appear that there is a new opportunity to develop realistically complex models that provide more reliable outputs and that can complete in a timescale suitable for a rich and interactive online, in addition to batch, approach.

Model development is a complex process that requires the collaboration of researchers and practitioners that span a range of disciplines, including statisticians, epidemiologists, health economists and health care managers. There are communities that have considerable experience in the development of computer systems that support this form of scientific co-operation. e-Labs are an emerging e-science paradigm that support the creation of social networks and provide generic applications that can be used to manage and visualise shared information [5]. Adoption of these frameworks provides significant advantages to both the users of healthcare systems and their developers. Users are presented with a familiar common interface for performing a range tasks that allow simple collaboration with colleagues. Generic services allow functionality to be shared across different tasks and avoid duplication of effort.

This paper presents an informatics approach to facilitate interoperability of the IMPACT system with emergent e-Lab frameworks, with the aim of supporting the social aspects of model development and sharing. It discusses the development of domain independent data models for this purpose and examines their usefulness for agile software development. It presents the results of the application of functional programming paradigm to the implementation of simulation software, supporting parallelisation and the use of capacity compute infrastructures.

**Data models**

In order to share a simulation through an e-Lab infrastructure, sufficient information must be passed between the simulator and e-Lab to completely characterise that simulation for the purposes of re-use and interpretation. Simulation data will include: the graphical disease model, the interventions and their associated relative risks and uptakes, the simulated cohort and the results. This will need to be supported with information that details the applicability of the data. For example, geographical variation in demographics will mean that the simulation results only apply to those regions that are well represented by the simulated cohort.

\(^3\) http://www.healthimpact.org.uk
It is not possible to interpret simulation data without contextual information, including the sources of data, such as databases and publications. This provenance is a requirement of the scientific process, being needed for citation and attribution. Contextual information will include details of the statistical methods along with their implementation.

Interpretation and re-use of the data are not possible without semantics at the data element level. Guidance is taken from ISO/IEC 11179-3, which outlines the basic attributes required to completely characterize a data element. These include: units, data-types, permissible values and precision. The system also needs to ensure that the concepts that relate to the data elements are well defined.

Sharing also requires standard models for representing simulation data. The generic nature of e-Lab services means that a domain independent data model must be used. IMPACT is considering the use of the Resource Description Framework (RDF) for this purpose. RDF has an abstract syntax that reflects a graph-based data model. It is simple, flexible and can be effectively used to express complex relationships between data. It has seen widespread use as the data model adopted by the Semantic Web.

The models used to represent data held within the IMPACT system are not constrained by those models used to share that data. For example, the simulator could store data in a relation database, with a schema that is tightly bound to the simulation domain. When the data is to be shared, a transformation can be made to the exchange format. The IMPACT software is developed using an agile paradigm with short development cycles and close interaction between software engineers, epidemiologists and statisticians. The nature of data models in an agile environment is an important and often under discussed consideration. With Waterfall software development, data models can be designed with full knowledge of the data requirements. The model can be very specific to the system and optimized assuming no changes. This is difficult with agile development, requirements may change and the data model must be designed with this in mind. Adoption of a domain independent model for the storage of data by the simulator would provide this flexibility for change. It is reasonable therefore, to consider the adoption of the Resource Description Framework for this purpose. Transformation for exchange is then a simple process. It is important to note that IMPACT users do not need to be aware of the underlying data standards and tools employed. When the data is managed through the system, the interfaces will be designed to abstract away such concerns.

Users may need to export data from the system for analysis elsewhere. The system will always allow export of the underlying RDF using a serialization standard such as N3 or Turtle. However, these representations are verbose and transformation would be required before data could be used with most statistical modelling packages, such as R. For practical purposes, the simulator will allow export in tabular formats that are easier to import.

**e-Labs and Work Objects**

The concepts of an e-Lab and Work Objects have been described elsewhere [5]. An e-Lab is described as ‘An information system for bringing together people, data and analytical methods at the point of investigation or decision making’. These characteristics closely match those desired for the IMPACT system. An e-Lab framework is being developed as part of the North West e-Health (NWeH) project [6] and IMPACT plans to make use of this framework to fulfil these objectives. The Work Object is a key concept for the e-Lab: they are aggregations of resources that support the sharing of
information. A resource is anything that can be part of a collection – abstract or concrete. As an example, a Simulation Work Object might consist of a publication, disease model, population, medical interventions, simulation results and users of the system. At the very least, all e-Lab generic services can work with these aggregations. They can be visualised, shared and persisted.

It is argued elsewhere that existing best practice for exposing, sharing and connecting data on the web, such as Open Linked Data, require a standard model for expressing aggregations of resources and the relationships between those resources [7]. The Research Object Upper Model (ROUM) is being developed by the e-Lab Technical Architecture Group (a collaboration of researchers at the University of Manchester and University of Southampton) for this purpose. The ROUM builds on OAI-ORE to provide a model that can be used to describe a basic aggregation and to provide additional semantics about the resources and relationships between them. The ROUM can be considered to specify the basic model that should be understood by all e-Lab systems. Simulator data that is to be used within an e-Lab must make use of this model to include a description of the aggregation that is the Simulation Work Object. It is a principal of the e-Lab system that it should support ‘graceful degradation of understanding’. Generic services should be able to consume Work Objects without understanding all of the content.

Most users will need to do more than just work at the level of the aggregation and its resources. For example, users may want to perform some analysis that might depend on some domain specific data. These applications will require a domain specific model that will need to be specified outside of the upper model. These domain extensions are known as Research Object Domain Schema (RODS).

**Simulation Work Objects**

IMPACT is developing a model that can be used to describe a Simulation Work Object. Each object aggregates all of the resources that relate to a single simulation. This includes: people, statistical methods, source publications, models, simulations, results, analysis, and output publications. The IMPACT RODS is being developed as an ontology to provide consistency with the use of other Semantic Web approaches and to open up the future possibility of inference over the simulation data. The ontology is specified using the Web Ontology Language (OWL) and is developed using the Protégé ontology editor. The ontology builds on the ROUM and other existing models, for example the Open Provenance Model. The IMPACT ontology is used to explicitly represent the data concepts and the relationships between those concepts. It is a conceptual model that provides the vocabulary used in the RDF representation of the data, providing a way to formally denote the data semantics. However, this model does not place constraints on the information content of the Work Object description. Any restrictions (such as cardinality) are used for the purposes of interference and the adoption of the open world assumption means that missing information does not produce an inconsistency. Therefore, it is proposed here that a Work Object model should consist of both a conceptual model and logical map model. The map model adopts closed world semantics and can be used to validate the content of a specific information source. The use of SPARQL queries to specify these constraints is under investigation. Figure 1 details the relationships between Work Object models.
Simulation Object Archives

It is important to emphasize that a Work Object is an abstract entity. It is the description of the Work Object that has a concrete representation that can be exported from the system. This distinction is made in OAI-ORE, where an Aggregation is abstract and described by a ResourceMap. Various serialization standards are put forward for the description, including RDF/XML, Atom and RDFa. IMPACT will allow users to export descriptions of Work Objects packaged with representations of the aggregated resources. Initially, these archives will be implemented as ZIP files. RDF/XML will be used to serialize the descriptions and to represent the aggregated resources. Other serialization standards will be examined in terms of their impact on parsing performance and size. Add something here about locally relevant data and

Figure 1. Work Object Models

Figure 2. Simulation Work Object archives
**Graph persistence**

This paper describes the use of domain independent models for both the exchange of Work Object descriptions and for the internal persistence of data by the simulator software. The use of RDF for this purpose opens up the possibility of using existing triple stores for this data. The increased use of RDF has seen a growth in the number of storage systems and their performance, both in terms of query speed and capacity. For these purposes, there are a number of criteria that such a system must meet. Firstly, the current version of the simulator is written using the .NET framework and any storage system must be accessible from .NET languages. Semantics Server from Intellidimension is one of the only commercial offerings, based on SQL Server. However, there many open source, non-commercial options available. Connectors such as SemWeb and dotNetRDF can be used to connect to third party tools such as AllegroGraph, 4store, Joseki, Sesame, Talis and Virtuoso. It is also possible to make use of some Java stores including Jena.NET through IKVM. Other stores, such as Sesame have REST based interfaces that can be accessed over HTTP. Another promising option includes Zentity, a research output repository platform developed by Microsoft Research. It is a hybrid store (a mixture of a relational database and triple store) that employs a domain independent data model, allowing data to be stored in a semantically rich way.

The existing IMPACT system uses a single data management component that provides all the data storage, retrieval and validation services for the other system components. This ensures that all of the business logic is localised to a single piece of software, therefore reducing the likelihood of inconsistencies across the system. The data management component makes use of the NHibernate framework to map a domain object model to a relational schema and stores the data using SQL Server. The use of a mapping tool provides an abstraction that allows the system developers to work with the domain model and not to be concerned with the details of the storage device. It will be useful to maintain this abstraction when the data is stored as RDF. Domain objects need to be mapped to RDF to facilitate all CRUD operations. The Hibernate API is widely used and accepted, making it a good choice for a new object-RDF mapping tool. The Topaz project had similar objectives but it developed a mapping tool for Java objects.

IMPACT is considering the development of a new object-RDF mapping tool that will be usable from .NET languages and that will support CRUD operations on domain objects. It will follow the familiar Hibernate API and be annotation driven to allow simple and intuitive use by system developers. As described earlier, the main purpose of using the RDF as a model for persisting data within the simulator is to support the agile development of the system. It is proposed here that such support could be provided using the concept of Named Graphs. This provides a route to reification, a way to make statements about statements. It is proposed that the OO-RDF mapping tool create separate RDF graphs for each release of the software. Different versions of the software might have different business logic and RDF statements from all versions of the system may not be valid.

**Functional Programming**

The IMPACT system employs two different models to simulate policy decisions: population and clinical. The population model is an accelerated failure time model (known as the Stock of Health) that allows the time of disease incidence to be determined for all members of a cohort. The clinical

---

4 http://www.topazproject.org/trac/wiki/Topaz
model forms the basis of a discrete event simulation, and is used to determine the disease history of members of a cohort. Both simulations employ a stochastic approach and large cohorts need to be simulated before a comparison can be made with observed data (such as mortality rates). Details of the simulation models are described elsewhere [8]. The requirement to provide an online interface to configure, execute and analyse simulations and the computationally expensive nature of the approach requires careful consideration of the software design and implementation.

Increasing, software executes on multi-core systems, ranging from desktop systems to IBM Blue Gene. Multi-threaded software must be developed to take advantage of these new machines, something that needs to be considered from the start of the development. Adoption of a functional programming paradigm is a natural choice for multi-threaded code. It encourages practices leading to software that can be easily parallelised or even automatically parallelised. This derives in part from the way that state is handled. A functional approach favours immutability and a lack of state in contrast to an imperative approach. IMPACT is developing simulation software using F#, a functional language that has full support in .NET 4.0. Developers are able to make use of a new asynchronous programming model to develop parallel code without having to handle the thread pool, locks or the propagation of exceptions. F# also provides support for static checking and inference of units-of-measure, helping to avoid the misuse to data passed between different functions. There are some overheads in the adoption of functional languages, particular because many developers have more experience working with an imperative approach. F# does provide some imperative constructs, such as if conditions and for and while loops. Trivial parallelisation of a simulation is afforded by breaking the cohort into a series of sub-cohorts that are simulated separately. Each member of a sub-cohort will be simulated in series using the same processor core. The optimal number of sub-cohorts must be chosen to be suitable for the particular architecture, preventing unnecessary and expensive context switching. The simulation methods are chosen to avoid any coupling between simulated individuals. This does have some implications for the model application, for example it is difficult to model situations where resource constraints are an issue. However, it does mean that little information has to be shared between threads during the simulation, making a multi-process approach very attractive. The computation could be shared across machine boundaries, allowing the use of high throughput resources including those offered by Cloud and Grid computing. The Stock of Health simulation software has been developed using these principles and offers a promising level of performance. An online recursive approach is employed to generate the required summary statistics, reducing the memory footprint and enabling the simulation of very large cohorts (typically 60 million individuals). With computation across two Intel Xeon E5520 processors with 34GB of shared RAM, the simulation time is reduced by two orders of magnitude when compared with an R-based prototype.

Conclusion

This paper has presented new concepts being employed to enable the IMPACT healthcare policy simulator to interoperate with emerging e-Lab frameworks for social networking and scientific cooperation. Detailed consideration is made of the informatics approach, building on the Work Object concept and making use of the modelling approach described for Research Objects. It outlines the usefulness of domain independent models, not just for the exchange of information between the simulator and e-Labs, but also for the persistence of data in an agile development environment. A case is made for the development of a simple annotation driven object-RDF mapping tool for the
.NET framework. A functional paradigm has been adopted for the development of simulation software that has illustrated a reduction in software development time coupled with significant performance benefits.

Acknowledgements

This work was supported by funding from the UK NIHR Greater Manchester CLAHRC programme. Thanks go to the e-Lab Technical Architecture group for many insightful discussions.

References


Distribution

Paper read at IEEE Computer Based Medical Systems 2011.
