Conference Report

Addendum to Informatics for Health 2017: advancing both science and practice

Philip J. Scott, *Centre for Healthcare Modelling & Informatics, University of Portsmouth, Portsmouth, UK*

Ronald Cornet, *Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands and Linköping University, Linköping, Sweden*

Colin McCowan, *Robertson Centre for Biostatistics, University of Glasgow, Glasgow, UK*

Niels Peek, *Health e-Research Centre, Division of Informatics, Imaging and Data Science, University of Manchester, Manchester, UK*

INTRODUCTION

This article presents presentation and poster abstracts that were mistakenly omitted from the original publication.[1]

[1] Scott PJ et al. Informatics for Health 2017: advancing both science and practice. *Journal of Innovation in Health Informatics* 2017;24(1):1–185.

SECTION 1: PRESENTATION ABSTRACTS

Abstract no. 268   Meeting clinicians' needs in the design of a Personal Health Record

Ian Tite, *Hitachi Consulting, London, UK*

Henry Potts, *UCL Institute of Health Informatics, London, UK*

Introduction   Central to the digital strategy of the UK's National Health Service and that of other health services around the world is the Personal Health Record (PHR), where individuals have access to and some control over their own medical record. The endpoint for PHR development is often seen as the individual becoming the custodian of their own medical data, with clinicians contributing to the data at each care episode. However, medical data is traditionally maintained by healthcare professionals, with the individual rarely directly contributing. How an individual-controlled PHR might work in the UK context needs investigation?

Method   We used a formal method to develop an information architecture, The Open Group's Architecture Framework (TOGAF). TOGAF helps demystify the process of development and provides reference to common elements expected in contemporary architectures. Without the use of a formal method, users' needs can remain poorly understood and designs may rest on flawed assumptions. An initial architecture reference model for a PHR was developed based on the literature. This was then used to guide a series of semi-structured interviews with five clinicians to elicit requirements. TOGAF was used as a guide to populate a proposed architecture for a patient-controlled PHR meeting clinicians' needs, including design principles addressing specification, a conceptual model and a list of architecture elements to be considered.

Discussion   Identified features included: patient access to the medical record identity management, high levels of information security, secure messaging between patient and clinician, electronic consultations, robust audit trails, accessibility at the point of care device, independence resilience and availability to high standards of Service Level Agreement consent management and standard messaging (e.g. Health Level 7 (HL7), HL7 Fast Healthcare Interoperability Resources (FHIR)) integration with clinical systems and limited decision support. These requirements could be accommodated with some development by contemporary PHR technologies. However, a central question emerged about how much an individual should steward of their own data. Participants all voiced the need for integrity of the record via the control of access, with editing rights being granted only to authorised clinicians. The ability to filter parts of the medical record, especially clinicians' notes and episodic data such as test results in certain circumstances, was desired. Augmentation of the record by individuals was seen as acceptable.

Conclusion   A PHR that meets most clinicians' requirements can be built today. But desired design elements, for example, filtering of the medical record, directly contradict government policy, which promises online access to medical records by 2018, and contemporary views on the advantages of transparency in the patient/professional relationship. Such filtering may even contravene the Data Protection Act; however, an exception can be made where a clinician wishes to avoid distress, where viewing of data without the support of a clinician would be harmful. This area requires greater investigation to develop clear policies and architectures. An information architecture approach presents a useful lens on the requirements of clinicians if they are to successfully use a PHR.

Abstract no. 300   Community detection algorithms for analysis of biological networks

Jonathan Silva, *Department of Informatics, Faculty of Natural and Mathematical Sciences, King's College London, London, UK*

Laura Bennett, *Centre for Process Systems Engineering, Department of Chemical Engineering, University College London, London, UK*

Aristotelis Kittas, *Department of Informatics, Faculty of Natural and Mathematical Sciences, King's College London, London, UK*

Lingjian Yang, *Centre for Process Systems Engineering, Department of Chemical Engineering, University College London, London, UK*

Songsong Liu, *Centre for Process Systems Engineering, Department of Chemical Engineering, University College London, London, UK*

Lazaros Papageorgiou, *Centre for Process Systems Engineering, Department of Chemical Engineering, University College London, London, UK*

Sophia Tsoka, *Department of Informatics, Faculty of Natural and Mathematical Sciences, King's College London, London, UK*

Introduction   Community structure is often found in complex networks where the vertices are naturally clustered into tightly connected modules with a large number of within-module edges and a few inter-module links. The ability to identify and analyse such structures has proven to be important in revealing the underlying properties of these networks. For example, sub-groups in metabolic and cellular networks may reflect distinct functions in biological systems, and the evolutionary properties of biological molecules and species modules found in drug networks may help understand the mechanism of action of compounds where this process is still unknown and help re-purpose the existing drugs to treat different diseases. Therefore, the modular view of networks provides a clearer understanding on how complex systems are constructed from a number of fundamental components and sheds light into the interactions of such components.

Method   We present a suite of optimisation models for detecting communities by maximising modularity, a popular metric of the goodness of network partitions, reformulated to each type of problem accordingly. We implement these algorithms using mathematical programming techniques, where optimisation problems are defined by a set of specific mathematical rules expressed as constraints and objective functions and later solved by suitable solution algorithms, commonly based on branch-and-bound principles. This framework has been shown to outperform state-of-the-art algorithms, also allowing for more mathematically descriptive solutions to community detection in biological networks. Such methodologies can be easily modified to incorporate additional features according to specific user requirements, for example, constraining the minimum/maximum module size or balancing among modules, thereby enabling more flexible modelling frameworks. More specifically, we showcase algorithms designed for different types of networks: OptMod, for static undirected networks (weighted or unweighted); DiMod, for directed networks; OverMod, a model that detects overlapping modules, to account for the fact that entities in a network might belong to more than one community; SeqMod, an extension of previous models to detect groups in a temporal network; and finally, SimMod, for the type of networks with multiple layers, also called multi-layer or multiplex networks.

Discussion   A case is made for the application of these methods to biological and biomedical data. We show the application of these algorithms to detect groups in various interaction networks and application domains. In a particular case study, we

illustrate the integration of biological data from various sources to implement community detection via consensus clustering (SimMod). Such models that can accommodate multiple data inputs and produce composite communities spanning diverse interaction sets are thought to reflect biological knowledge more accurately and may lead to development of enhanced bio-medical applications.

Abstract no. 314   Which outcomes matter to patients? Comparing the relationship between patient-reported and traditional outcome measures on patient satisfaction in surgery

Catrin Jones, *Royal Infirmary of Edinburgh, Edinburgh, UK*

Thomas Drake, *Royal Infirmary of Edinburgh, Edinburgh, UK*

Stephen O'Neill, *Royal Infirmary of Edinburgh, Edinburgh, UK*

Kenneth McLean, *Royal Infirmary of Edinburgh, Edinburgh, UK*

Catherine Shaw, *Royal Infirmary of Edinburgh, Edinburgh, UK*

Stephen Wigmore, *Royal Infirmary of Edinburgh, Edinburgh, UK*

Ewen Harrison, *Royal Infirmary of Edinburgh, Edinburgh, UK*

Aims   Surgical outcomes have traditionally been described using clinician-centred measures such as 30-day mortality, complication and reintervention rates. However, there has been a growing recognition that these measures may fail to capture the outcomes that truly matter to patients and, therefore, are poorly equipped to drive improvement in patient-centred care. This study explored the relationship among patient experience measures, patient reported outcome measures and traditional outcome measures on the overall satisfaction of surgical patients with the outcome of their procedure.

Methods   The patient demographics, operation details and 30-day clinical outcome data of every patient who underwent emergency intra-abdominal surgery at a large teaching hospital over a 4-week period and patients who underwent cholecystectomy over an 8-week period were collected. Ethical approval was obtained to administer validated Patient Reported Experience Measures and Patient Reported Outcome Measures questionnaires. Multiple regression analysis was performed for factors significantly associated with satisfaction in univariate analysis.

Results   For patients undergoing emergency, intra-abdominal surgery, good communication (b = −1.26, (95%, CI: 0.18–2.34), p = 0.020), trust in staff (b = −1.51, (95%, CI: 0.54–2.49), p = 0.003) and a quiet ward (b = −1.35, (95%, CI: 0.56–2.14), p = 0.001) were all strongly associated with increased overall patient satisfaction on linear regression analysis. Minor complications (b = 0.3 and p = 0.405), reintervention (b = −0.4, p = 0.640), surgical sites infection (b = 0.5 and p = 0.586) or abscess (b = 1.5 and p = 0.298) had no effect. Similarly, for patients who had cholecystectomies, knowing who to contact (b0.17 (95%, CI: 0.05–0.29), p = 0.008) was associated with higher satisfaction, while complications (b0.26 (95%, CI: −0.07–0.59), p = 0.12) had no effect.

Conclusions   Traditional outcome measures are poor predictors for patient satisfaction and fail to capture factors that are most important to patients. Patient reported experience and outcome measures should be used more widely and given more importance in evaluating surgical outcomes.

Abstract no. 326   Clinical Decision Support for Diabetes in Scotland: evaluation of clinical processes and outcomes

Nicholas Conway, *University of Dundee, Dundee, UK*

Scott Cunningham, *NHS Tayside, Dundee, UK*

Ann Wales, *University of Dundee, Dundee, UK*

Deborah Wake, *NHS Education for Scotland, Glasgow, UK*

Introduction   Nearly three million people in the UK have diabetes (>6% of adults), with prevalence expected to double over the next two decades. Clinician Decision Support Systems (CDSS) are associated with improved adherence to clinical guidelines. CDSS was implemented within the Scottish National Electronic Health Record, Scottish Care Information (SCI) Diabetes and has been live within NHS Tayside and Lothian since December 2013, serving a combined diabetes population of ~30,000. This study aims to describe users' attitudes and reactions to the system and to quantify impact on clinical processes and outcomes.

Method   Health care professional (HCP) opinion was sought via focus groups and questionnaires. SCI-Diabetes data were extracted for two time periods: December 2013–February 2014 (Ninewells Hospital, Dundee, UK) and August 2014–November 2014 (St John's Hospital, Livingston, UK). SCI-Diabetes usage was quantified using HCP interaction ('mouse-clicks') and time spent within the patient record. HCP behaviour was compared between instances where CDSS messages were displayed, with instances where not (corrected for user-role, patient age, diabetes type/duration, co-morbidity and deprivation). Case-control comparison was made to assess clinical processes and outcomes. Cases were patients whose HCP received a CDSS message during the consultation. Controls were matched for age, sex, diabetes type and duration BMI and clinic attendance in areas outwith the pilot. Clinical process measures were screening for hypercholesterolaemia, kidney, foot and thyroid disease. Clinical outcomes included HbA1c cholesterol blood pressure and urinary albumin/creatinine (UACR) at 1 year. Comparison was made using multivariable regression.

Results   Pre- and post-intervention HCP questionnaire response rates were 57/105 (54%) and 39/105 (37%). Three focus groups were held (n = 8–9/group). The majority of respondents/participants had a positive or neutral response to the system. Early-adopters reported usage within clinical workflow. CDSS messages were displayed on opening 6,665/17,280 (39%) records. For nurses, the presence of CDSS message was associated with increased SCI-Diabetes usage, compared with instances where no message was displayed (median 'clicks' 19 (IQrange: 8–37) versus 16 (7–32), adj.p = 0.014). For doctors, CDSS messages were associated with reduced time within the patient record (median duration 33 (IQrange: 5–86) versus 38 (12–97) secs, adj.p = 0.032). 1,883 cases attended clinic, matched to 3,557 controls. Probability of receiving screening more than doubled for hypercholesterolaemia (adjOR 2.4, (95%, CI: 1.6–3.0)), creatinine (2.5 (1.6–3.9)), UACR (2.3 (1.9–2.8)) and foot screening (2.9 (2.3–3.6)) – all p < 0.001. Screening for hypothyroidism increased slightly (0.8 (0.7–1.0), p = 0.035). For those attending clinic, study group did not predict clinical outcomes at 1 year. Post-hoc analysis of all patients with a CDSS prompt (n = 5,692) showed small improvements in mean HbA1c (−2.3mmol/mol versus −1.1, B1.2 (0.4–2.0), p = 0.003) compared to controls (n = 10,667). Mean UACR increased in both groups but more so in controls (baseline 8.7 mg/mmol versus 9.3, increasing by +1.6 versus +4.4, B2.9 (0.7–5.1), p = 0.01).

Discussion and Conclusion   The CDSS was associated with improved efficiencies in working practices (dependent on role) and large improvements in guideline adherence. If replicated nationally, thousands more individuals would receive appropriate screening tests. These evidence-based early interventions can significantly impact on costly and devastating complications such as foot ulcers, amputations, cardiovascular disease, renal failure and death. The potential benefits of this project extend beyond the Scottish diabetes population, as NHS Scotland considers how best to realise the full potential of CDSS.

Abstract no. 330   Evaluation of CPRD GOLD e-learning

Yuewen Wu, *Clinical Practice Research Datalink, London, UK*

Yvonne Kaserebe, *Clinical Practice Research Datalink, London, UK*

Helen Strongman, *Clinical Practice Research Datalink, London, UK*

Rachael Williams, *Clinical Practice Research Datalink, London, UK*

Introduction   Clinical Practice Research Datalink (CPRD) provides a one-day training course to introduce users to the CPRD real-world primary care database known as "GOLD". Historically, the course has been delivered face to face or through video conferencing. Demand has grown due to an increase in organisations and individuals who use CPRD data. In March 2015, CPRD replaced this course with a Massive Open Online Course (MOOC). The MOOC is available through the Medicines and Healthcare Products Regulatory Agency learning portal, free of charge, to anyone with an interest in CPRD. It was created to build knowledge about CPRD GOLD data and how to use it for health research and enable resources to be re-allocated to other priority work streams. In this study, we set out to evaluate the effectiveness of the CPRD GOLD e-learning course in meeting these objectives.

Method   A descriptive quantitative analysis of user audit data from the e-learning tracking system was carried out to evaluate the uptake and completion of the CPRD GOLD e-learning. Qualitative e-mail and telephone surveys were carried out with seven completers and three non-completers to assess the effectiveness of the course as a learning resource.

Results   From March 2015 to October 2016, 390 people registered on the e-learning system of whom 250 completed at least one module and 161 completed all modules. This compares to 132 individuals who were trained through face-to-face or real-time online delivery over the preceding 18 months. Significant researcher time was invested in developing the e-learning and there has been considerable learning along the way. However, this has since then freed up 10–20 days of CPRD trainer time per annum. For the nominated users from a CPRD GOLD licensee organisation, it is mandatory to complete all the e-learning modules and 24 of the 161 completers were nominated users. The total distinct count of registered organisations was 74, among which 46 had at least one completer. The 46 organisations included 22 academic, 18 pharmaceutical, 2 government and 4 other organisations. All seven of the completers improved knowledge about CPRD from the e-learning and six considered the course to have been useful for their work. Recommendations for improvement included providing real-world case studies and an online forum for questions. Reasons for not completing related to early modules satisfying learning requirements when the course was accessed.

Discussion   The CPRD e-learning has reached a larger audience than the traditional training method, at a lower ongoing cost to resource. The interest in the e-learning and CPRD is beyond CPRD GOLD nominated users and licensee organisations. Qualitative feedback was highly positive and provided useful suggestions for improvement.

Conclusion   E-learning has attracted a wide interest and allowed CPRD to build capacity of a wider audience, contributing to CPRD's mission of supporting public health research. Experience from developing the CPRD GOLD e-learning has also been used to build a new e-learning course aimed at general practitioners (GPs) who are interested in real-world pragmatic studies supported by CPRD.

Abstract no. 338   Evaluating the impacts on health outcomes of Welsh Government funded schemes designed to improve the energy efficiency of the homes of low-income households

Sian Morrison-Rees, *ADRC, Swansea University, Swansea, UK*

Introduction   Living in fuel poverty often means living in an inadequately heated house. The World Health Organisation (2007) recognises that living in a cold and/or damp house may be harmful to health, increasing the risk of morbidity, mortality and excess winter deaths. As part of its strategy to reduce fuel poverty in Wales, the Welsh Government developed a demand-led fuel poverty scheme called Nest to improve the energy efficiency of homes. Warm Homes Nest provided home energy efficiency improvements to those most likely affected by fuel poverty, including low income and vulnerable households from 2011 to 2015. The energy efficiency measures provided included insulation and heating upgrades, such as a more efficient boiler. The overall aim of the project is to evaluate the health impacts of Welsh Government funded schemes through the use of existing data linked to the routine health records held in the SAIL Databank at Swansea University.

Methods   A longitudinal data set was created using the anonymised residential dwelling that has received home energy efficiency improvements linked to a summary of their health measures (hospital admissions, reason for the admission, GP prescriptions and clinical diagnoses).

For each year of the study period, we used a stepped wedge design to construct cohorts of people who had already received the intervention and for a control group of people who had applied for measures but not yet received them. We used difference in difference estimations to compare any changes in the health of people before and after the intervention with any concurrent change in health in those who required, but had yet to receive, the intervention. Our first analysis focuses on cardiovascular, respiratory and general health.

Results   The anonymised linking process created a data set for over 35,000 individuals of all ages living in homes that received home energy efficiency upgrades. An early and indicative analysis of the data suggests a positive impact on health for recipients of the Nest scheme. Recipients had decreased the rates of hospital admissions for both respiratory and cardiovascular conditions the winter after measures were installed compared to those who were eligible for the Nest scheme but who had not yet received measures. Recipients of the Nest scheme also had a smaller increase in GP prescriptions the following winter than those waiting for measures, suggesting a 'protective effect' on overall general health. We anticipate concluding the analysis in 2016 in order to inform the development of a future Welsh Government demand-led fuel poverty scheme due to succeed Nest from September 2017. Our results will compare specific interventions for their impacts on health. We will be able to report whether particular population groups, for example, those suffering from particular health conditions gain particular benefit from interventions.

Discussion   Providing home energy efficiency interventions has the potential to benefit population health. Our findings will inform more effectively focussed home energy efficiency schemes in order to reduce the numbers of people living in fuel poverty and, thus, improve the health and wellbeing of people living in Wales.

Abstract no. 350   Involving physical activity in insulin recommender systems with the use of wearables

Beatriz López, *University of Girona, Girona, Spain*

Alejandro Pozo, *University of Girona, Girona, Spain*

Ferran Torrent-Fontbona, *University of Girona, Girona, Spain*

Introduction   Type Diabetes Mellitus 1 (TDM1 patients are able to determine the amount of insulin to be injected in a dose according to the most recent food ingested and other factors such as physical activity or menstruation. Dealing with all these factors is a complex task, and patients suffering this illness are very active in looking for tools that can help them in these daily decisions. In that regard, insulin recommender systems are decision support system (DSS) designed with the aim of providing the appropriate insulin dose to a given patient in a given moment. Moreover, the deployment of such kind of DSS in mobile devices is offering the opportunity to use new sensors that may provide additional information to improve the recommendations. For example, sensors like smartwatches or wrist bands offer the opportunity to track patients' physical activity or even their stress level in order to feed the next insulin recommendation decision with this information. Our work concerns the development of an adaptive recommender system that exploits the information from wearables, in order to improve the recommendation provided to TDM1 patients.

Method   The recommender system relies on case-based reasoning methodology. This methodology has been proved to be useful in medical domains, since it is able to provide personalised recommendation to patients. The recommendations are based on several parameters such as carbohydrates and fats ingested, recent and future physical activity, stress, etc. In that regard, the information entered manually by people regarding physical activity is subjective, as it depends on the people's appreciation about that concept. Walking 500 m could not be an exercise if it is performed with calm, but it could be if walked at a given pace. On the other hand, smartwatches and wristbands are wearable sensors able to provide some measures from which to estimate the type and intensity of the physical activity. In that regard, the estimation depends on the average physical activity of the user and is unique (personalised) for each person. Our work consists of a physical activity module that complements the recommender system with the gathered and processed sensor data. It reads the steps data provided by a sensor and returns the type and intensity of the activity performed before a bolus recommendation. Two types of physical activity are being considered (aerobic and nonaerobic) and four levels of intensity. The quantified intensity depends on the average physical activity of the user.

Results   The system has been implemented using the eXiTCBR tool. The first prototype will be tested in 2017.

Discussion   Physical activity quantification removes uncertainty involved in the person's subjective evaluation about that concept. The system requires some initial data. Other consideration to be taken into account, in order to exploit mobile devices, is contextual information.

Conclusion   Providing useful insulin recommendations requires the management of several parameters, among them the physical activity. Physical activity is a subjective concept. Wearables offer the opportunity for estimating the physical activity. This project has received funding from Horizon 2020 No. 689810.

Abstract no. 393   Supporting biomarker discovery using text mining

Paul Thompson, *University of Manchester, Manchester, UK*

Katherine Boylan, *University of Manchester, Manchester, UK*

Anthony Freemont, *University of Manchester, Manchester, UK*

Sophia Ananiadou, *University of Manchester, Manchester, UK*

Introduction   The Manchester Molecular Pathology Innovation Centre (MMPathIC) is creating an environment to develop new biomarker tests, using molecular pathology techniques to facilitate patient stratification. To ensure well-informed decisions, MMPathIC combines medical expertise with skills from other research areas. In particular, text mining (TM) techniques are being applied to vast volumes of unstructured electronic text, to automatically locate and link various types of biomarker-related information, which may remain hidden using traditional search techniques.

Methods   We are using TM techniques to detect various aspects of the semantic structure of text, for example, recognition of concept mentions (genes, their variants, diseases, risk factors, drugs, patient groups, etc.) and relationships among these concepts (variants of a gene having an association with a disease in specific types of patients, etc). The Argo web-based TM workbench (http://argo.nactem.ac.uk/) facilitates complex processing of text, by integrating various TM tools and machine learning capabilities, allowing tools to be tailored to specific tasks. Tools are combined into TM processing pipelines, performing various levels of linguistic and semantic processing to recognise complex information in large document collections.

Results   While in the early stages of the project, we are developing and applying Argo pipelines to a sub-set of MEDLINE abstracts to assess their performance. First, we are combining the outputs of several concept recognition tools, taking advantage of their differing strengths. We are also exploiting sentence structure to collect a set of linguistic patterns that are used to describe known gene-disease relations and applying these patterns to larger data sets to uncover novel associations. Subsequently, we will link in contextual information (such as patient characteristics, response to drugs, etc.) to create more complex relationships.

Discussion   We take inspiration from, but build upon, techniques that have been employed in existing systems that allow searching for gene-disease associations, for example, FACTA+ (http://www.nactem.ac.uk/facta-visualizer/) and DisGeNET (http://www.disgenet.org/). However, our approach will include a number of novel aspects, aimed at making it easier to discover and filter information of interest. These will include the detection of contextual information and linking of information that is dispersed across multiple documents, in order to construct more detailed and complex relationships amongst concepts. We will additionally classify the relations in different ways (e.g. according to whether a biomarker has diagnostic or prognostic value, whether the relation is stated as a hypothesis or an experimental result, if there is any degree of any speculation specified). Accordingly, it will be possible to locate answers to complex queries, for example, *In which population sub-groups is there evidence that Gene X is a putative biomarker for Disease Y?*

Conclusion   In contrast to many related efforts, Argo's cloud-based processing capabilities make it feasible to apply our TM pipelines not only to abstracts, but also to huge collections of full articles, which are likely to contain much richer information relating to biomarkers. Our ultimate aim is develop an advanced semantically oriented search environment that provides medical experts with the means to efficiently locate evidence to support or motivate the development of biomarker tests.

SECTION 2: POSTER ABSTRACTS

Abstract no. 31   A new data opportunity for community nutrition surveillance: estimating spatial patterning of dietary behaviours using grocery transaction data

Hiroshi Mamiya, *Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, QC, Canada*

Erica Moodie, *Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, QC, Canada*

David L. Buckeridge, *McGill Clinical and Health Informatics, Department of Epidemiology and Biostatistics, McGill University, Montreal, QC, Canada*

Objective   To demonstrate the application of digital grocery transaction records to measure community food selection with high spatial granularity.

Introduction   Unhealthy diet is becoming the most important preventable cause of chronic disease burden. Identification of neighbourhood-level inequality in healthy food selection is necessary for the planning of targeted and tailored community intervention. Although survey has been a primary public health tool to collect diet-related risk factors, cost of it prevents mass administration required to assess food selection at high geographic resolution. Marketing companies such as the Nielsen Corporation continuously collect and centralise scanned grocery transaction records from a geographically representative sample of retail food outlets to guide and evaluate product promotions. These data can be harnessed by public health researchers to develop a model for the demand of specific food(s) using store and neighbourhood attributes, providing a rich and detailed picture of the neighbourhood dietary preference. In this study, we generated a spatial profile of food selection from estimated sales in food outlets in Montreal, QC, Canada, using regular carbonated soft drinks (i.e. non-diet soda) as an initial example.

Methods   From the Nielsen Corporation, we obtained weekly grocery transaction data generated by a sample of 86 grocery stores and 42 pharmacies in Montreal in 2012. Extracted store-specific soda sales were standardised to a single serving size (240 ml) and averaged across 52 weeks, resulting in 128 data points. Using linear regression, natural log-transformed soda sales were modelled as a function of store type (grocery versus pharmacies), chain identification code and socio-demographic attributes of store neighbourhood (e.g. income and education) as measured by the 2011 Canadian Household Survey. The final model selected by cross-validation was applied to all operating chain grocery stores and pharmacies in 2012 (n = 980) recorded in a comprehensive and commonly available business establishment database. The predicted store-specific weekly average soda sales were divided by population denominator and spatially interpolated to provide a geographical representation of unhealthy food preference.

Conclusions   The current lack of neighbourhood-level dietary surveillance impedes effective public health actions aimed at encouraging healthy food selection, such as development of community health campaign and evaluation of neighbourhood-level response to social and economic policies, including taxation and food subsidiary program. Our method leverages existing grocery transaction data and stores location information to address the gap in population monitoring of nutrition status and urban foodscapes. Future applications of our methodology include demand prediction from other store types (e.g. convenience stores) and food products across multiple time points (e.g. mouths and years).

Abstract no. 56   C3-Cloud: a federated collaborative care and cure cloud architecture for addressing the needs of multi-morbidity and managing poly-pharmacy

Gökce Banu Laleci Erturkmen, *SRDC Software Research Development and Consultancy Corp., Ankara, Turkey*

Mustafa Yuksel, *SRDC Software Research Development and Consultancy Corp., Ankara, Turkey*

Theodoros Arvanitis, *Institute of Digital Healthcare, WMG, University of Warwick, Coventry, UK*

Introduction   A growing share of the population (15% in 2010) in Organization for Economic Cooperation and Development (OECD) countries is over 65 and expected to reach 22% by 2030. Older age is associated with an increased accumulation of multiple chronic conditions. More than half of all older people have at least three chronic conditions and a significant proportion has five or more. The clinical management of patients with multi-morbidity is much more complex, disconnected and time-consuming than that of those with single diseases. As a result, multi-morbid patients with long-term care need experience shortcomings and gaps in their care provision. There is an increasing need to organise the care around the patient with the involvement of all stakeholders, and as a response to this requirement, the C3-Cloud project aims to achieve high quality integrated care with the support of information and communication technologies (ICT).

Method   C3-Cloud will establish an ICT infrastructure to enable continuous coordination of patient-centred care activities by a multidisciplinary care team (MDT) and patients/informal care givers. A Personalised Care Plan Development Platform will allow, for the first time, collaborative creation and execution of personalised care plans for multi-morbid patients through systematic and semi-automatic reconciliation of clinical guidelines. This will be accomplished through Clinical Decision Support Modules for risk prediction and stratification, recommendation reconciliation, poly-pharmacy management and goal setting. Fusion of multimodal patient data will be achieved via C3-Cloud Interoperability Middleware for seamless integration with existing health/social care information systems. Active patient involvement and treatment adherence will be realised through a Patient Empowerment Platform, ensuring patient needs are respected in decision making. In order to demonstrate the feasibility of the C3-Cloud integrated care approach, pilot studies will focus on diabetes, heart failure, renal failure and depression in different comorbidity combinations and operate for 15 months in three European regions with diverse health and social care systems, in addition to a diverse ICT landscape (South Warwickshire, Basque Country, Jämtland-Härjedalen). In total, 150 patients for intense evaluation, 600 patients for large-scale impact assessment and 62 MDT members will be involved in pilot operation and evaluation activities.

Results   In the first five months of the project, ideal to-be scenarios have been produced by the end-users, which, following user-centred design principles, led to the identification of technical use cases and formal requirements of the C3-Cloud architecture. Work is now focused on the design of the architecture and the critical analysis of relevant clinical guidelines.

Discussion   Unfortunately, current European medical models focus primarily on short and medium term interventions on the basis of single conditions, failing to integrate care planning well across providers and often overlooking the interconnected basis of chronic diseases. Managing multi-morbidity, through the current treatment methods, results in specialty silos and fragmented care, involving multiple care providers who are not effectively sharing information.

Conclusion   As a result of 15 months of piloting in diverse settings, C3-Cloud aims to strengthen the evidence base in coordinated care of patients with multi-morbidity and to inform future development of more streamlined and optimised multi-morbidity care pathways.

Abstract no. 160   The Biomedical Informatics Network for Education, Research and Industry (BINERI) at the University of Leicester

Tim Beck, *University of Leicester, Leicester, UK*

Anthony Brookes, *University of Leicester, Leicester, UK*

Introduction   A wide range of biomedical informatics, healthcare data management and information technology expertise operate together in Leicester under a strategic grouping called Biomedical Informatics Network for Education, Research and Industry (BINERI). This network works in a unified way on topics such as bioinformatics training, data science, expertise sharing, data discovery and sharing, bio-banking, big data analysis, ethics, governance, patient engagement, information technology, etc.

Method   By integrating various complementary disciplines, combining resources, sharing Ph.D. students and staff, and applying jointly for local, national and international funding, BINERI members are propelling forward high impact informatics initiatives across its research and NHS Trust stakeholders. These stakeholders include the following.

- Leicester Precision Medicine Institute (LPMI)
- University Hospitals of Leicester NHS Trust
- Leicester Cardiovascular BRU
- Leicester Respiratory BRU
- Leicester-Loughborough Diet, Lifestyle and Physical Activity BRU
- Leicester Clinical Trials Unit
- NIHR Collaborative Leadership in Applied Health Research and Care
- Cancer Research UK Leicester Centre
- University of Leicester IT Services
- University of Leicester Bioinformatics and Biostatistics Analysis Support Hub
- University of Leicester College of Medicine, Biological Sciences and Psychology
- University of Leicester College of Science and Engineering.

   BINERI is founded on the principle that data (primary data, aggregated data, metadata and resulting knowledge) need to be used optimally – which implies aligning research and healthcare activities to ensure professional and forward-looking data capture, management, curation, security, discovery, sharing and re-use (as per Global Alliance for Genomics and Health (GA4GH) and "Findable, Accessible, Interoperable, and Reusable" (FAIR) principles). It will also be important to stress patient-centric aspects of such developments, not least human-computer interactions and patient control of 'their' data.

Results   BINERI teams promulgate their extensive experience in the required tooling, such as concerning healthcare data warehouses (e.g. i2b2, tranSMART and LabKey), tissue tracking (e.g. OpenSpecimen), patient relationship management (e.g. CiviCRM), electronic case report form platforms (e.g. REDCap (Research Electronic Data CAPture) and OpenClinica), Clinical Data Interchange Standards Consortium (CDISC) standards, data discovery (e.g. Cafe Variome) and sequence variant validation (e.g. VariantValidator).

Ongoing exemplar projects include the following.

- Integration of primary care data from more than 100 GP surgeries in the East Midlands for the purposes of follow-up cardiovascular and diabetes research.
- Pragmatic randomised controlled trial nested within UK primary care data to evaluate the real-life effectiveness of diabetes drugs.
- Technical underpinning of an EU-wide, federated network of Alzheimer's disease patient data sets enabling discovery (without revealing data) of subjects suitable for a multi-site longitudinal readiness cohort for clinical trials.
- Management of translational clinical and research data across large-scale UK and EU consortia in the context of precision medicine (e.g. respiratory disorders and radiotherapy).
- Development of technological platforms such as Apps for managing chronic conditions.

Discussion   The resulting vibrant and productive bioinformatics environment in Leicester is further strengthened by having a dedicated bioinformatics hub, launching and funding the LPMI, holding monthly workshops between data scientists, and initiating many cross-disciplinary studentships with an emphasis on bioinformatics and personalised medicine.

Conclusion   BINERI collaborations with teams across the UK and internationally are diverse and numerous, with the door very much open to discussing further opportunities.

Abstract no. 175   Simulated data: an object oriented approach

Athanasios Anastasiou, *Farr CIPHER – Swansea University Medical School, Swansea, UK*

The objective of this poster is to demonstrate the structure and use of DGen, an object-oriented approach to Data GENeration and DeGENeration of clinical data in various formats.

The generation of artificial data is a topical subject for clinical research as it offers ways to stress-test systems, test algorithms in the presence of known data and of course for training and educational purposes. The generation of artificial data is characterised by two competing specifications: data should be generated according to pre-specified rules and data should be realistic. The impact of these two competing specifications is high for educational purposes as, it is desirable to train students over known, simple test-cases but not so simple as to be obvious and 'text-book' examples.

A number of tools (and techniques) for the simulation of data in general and clinical data in particular already exist at various levels of complexity in installing, seting up and operating them. DGen attempts to address the problem of generating artificial clinical data, of realistic complexity, by a framework of elementary Data Generators and Pertubators. Generators are responsible for creating random variables with full control over the characteristics of their values, and Pertubators are responsible for applying commonly encountered errors such as punctuation, abbreviation, data omission and others to the generated values. DGen uses operator overloading to define a very simple 'algebra' of combining generators to form complex cases (such as conditionally probable ones) and generalisation to create more complex entities such as 'Patient', 'CasePatient', 'ControlPatient' and others.

DGen is written in Python and is by no means complete. Future work includes improving the way random variables are described and formalising the transformation to specific data formats via the use of renderers.

Abstract no. 204   User-oriented oncological wiki through requirements prioritisation based on Kano-classification

Danijela Schlue, *Niederrhein University of Applied Sciences, Krefeld, Germany*

Bernhard Breil, *Hochschule Niederrhein, Krefeld, Germany*

Jörg Haier, *Universitätsklinikum Hamburg-Eppendorf, Hamburg, Germany*

Introduction   To establish a national knowledge management system and a reference work for standardised cancer documentation, a prototype of an oncological wiki (http://www.tumor-wiki.de) was designed and implemented as a flexible model with different views. Usability and user friendliness of the existing web interface were tested in July and August 2016. The responses from six users varied considerably. The issue of the right requirements management was raised. The objective of this work is to involve the user in the requirement prioritization process, to determine specific options for actions, and to identify requirements for further research. The aim of this approach is to classify user requirements with regards to their effect on user satisfaction.

Methods   Requirements definitions are based on contact with project workers and analysis of scientific literature. The requirements data collections comprised an e-mail survey among medical experts and documentation officers in addition to telephone interviews. Informatics experts drafted a questionnaire based on the results of structured surveys with six categories: functions for compilation of terms, functions for term usage, data, ergonomics, interfaces and documentation. The Kano approach suggests a systematic process for classifying system features and associated user requirements based on a questionnaire. A set of 46 system features and requirements was identified. A questionnaire was created, which contains a set of question pairs

for each requirement. Questionnaires are presented to the future users, project partners from four German Comprehensive Cancer Centres and the group of medical documentarists.

Results   The questionnaire consists of 46 requirements in six categories on five pages. Three of ten asked persons (one physician and two documentation officers) sent completed questionnaires back. The most relevant category is the 'Must-be'-category, which currently contains 23 items. Seven items belong to the A-category. There were not any items in the Q-category.

Discussion   A first pre-test gave valuable input for further optimization to follow a user-oriented design. First, results of this questionnaire gave valuable insights as to which requirements are most important and which functions are critical for a broad user acceptance of the implemented wiki. A main disadvantage of the Kano-classification is the length of the questionnaire. The questions might also be misunderstandings while reading the requirements.

Abstract no. 366   Sharing the process: the performance and portability of the ICONIC de novo sequence assembly pipeline in a virtual environment

Anil Gunesh, *University College London, London, UK*

Dan Frampton, *University College London, London, UK*

Tiziano Gallo Cassarino, *University College London, London, UK*

Myrto Kremyda-Vlachou, *University College London, London, UK*

David Wong, *University College London, London, UK*

Andrew Hayward, *University College London, London, UK*

Eleni Nastouli, *University College Hospital, London, UK*

Paul Kellam, *Wellcome Trust Sanger Institute, Cambridge, UK*

Deenan Pillay, *Africa Health Research Institute, Durban, South Africa*

Zisis Kozlakidis, *University College London, London, UK*

Introduction   The rate and volume of data acquisition are ever increasing in the areas of clinical pathogen research especially with the advent of next generation sequencing and the link of genomic to healthcare records. However, in the majority of cases, downstream analysis of the sequencing data takes place in dedicated high-performance computing cluster research environments within Universities on physical CPUs and with a small potential of further portability. The current distributed, replicated NHS setup creates many unforeseen problems in analysing large sequencing data sets. Different NHS sites have different information technology systems, hardware and availability of human skills. Due to these differences, the installation and configuration of genomic analysis software is often time-consuming and complex and results in a non-standard, site-specific installation.

Methods   The ICONIC (InfeCtion respONse through vIrus genomiCs) project has created and successfully implemented a virus agnostic de novo genome assembly pipeline capable of generating whole genome sequences from raw sequencing reads for HIV (human immunodeficiency virus), HCV (hepatitis C virus), Influenza, MERS Co-V (Middle East Respiratory Syndrome coronavirus), Norovirus, RSV and Measles viruses. The purpose of the pipeline is to facilitate parallel processing of large numbers of viral samples within the current operations of the public health system: both routinely and in real time in response to an outbreak.

Results   The results of the pipeline have been used in the UCLH and Barts Health NHS Trusts to assemble viral genomes from clinical (residual diagnostic) samples and to report on genomic variability and underlying drug resistance. This information can be used by virologists applying appropriate treatments and/or infection control measures.

Virtualisation allows this software to be installed in a virtual environment in which the number of processors and allocated memory can be dynamically increased or decreased according to the demand. Virtualisation also allows the complex software pipeline to be easily ported and installed in a number of different virtual environments by greatly reducing the complexity of installation and configuration under local constraints. In this instance, virtualisation was achieved within the eMedlab environment using the Red hat Openstack Platform. Distribution of the pipeline has been achieved through Docker as the latter automates the deployment of applications within software containers.

Discussion   The aim of this work is to show that a virtual environment can easily accommodate the software pipeline and, when optimised, performance and portability are scalable. It demonstrates that larger volumes of data can be processed in

a shorter time, thereby reducing costs, variability and increasing consistency. The ICONIC project and the virtual appliance approach demonstrate how this method can be utilised as a practical solution to the current fragmented NHS environments as part of an established routine service.

Abstract no. 426   A process evaluation of a web-based self-management tool

Sara Ahmed, *McGill University, Montreal, QC, Canada*

Owis Eilayyan, *McGill University, Montreal, QC, Canada*

Marie-France Valois, *McGill University, Montreal, QC, Canada*

Introduction   My Asthma Portal (MAP) is a web-based self-management support system that couples evidence-based behavioural change components (self-monitoring of symptoms, physical activity and medication adherence) with real-time monitoring, feedback and support from a nurse case-manager. The objective of this study was to evaluate the process implementation of MAP over a 6-month period as part of a randomised controlled trial that evaluated the impact of MAP on asthma control and asthma-related quality of life.

Method   Forty-nine individuals randomised to MAP were asked to log-in a minimum of once a week over a period of 6 months. MAP provides: 1) tailored asthma education, 2) asthma-related personal health data and monitoring tools and 3) e-mail access to the care team with feedback to modify health behaviours related to medication adherence, action plan use and physical activity. Audit trails (computer logs) were used to evaluate the usage of MAP and the nurse case-management system. The Technology Acceptance Model was used to assess perceptions of usefulness and ease of use.

Results   The average time spent logged into MAP was 7 min/login (range 20 s to 2 h). Access to MAP during the first three months ranged 1–69 (mean, SD: 24, 16), and 0–38 between three and six months among the 47 individuals who received the intervention 16% had no logins between three and six months. Logins were more frequent in the first four weeks (mean, SD: 12, 8) and then tapered off thereafter (mean, SD: 36, 42). The most frequently accessed components of the portal were the monitoring questionnaire and the asthma feedback centre, with an average of 4.5 views in the first week, 2.2 in the second week, 1.5 from week 2 to 7 and then 1 time per week for the remainder of the trial. The least frequently accessed component was the static Learning Centre. Participants emailed the nurse to discuss medications (30%), clarify the action plan (18%), monitor MAP feedback (15%) and follow-up related to control status (5%). Sixty-one percent of alerts to the nurse were related to medication adherence, 17% indicated a need to review the action plan, 11% were because the patient did not login, and 7% were related to delayed initiation of the action plan by patients. Most participants (82%) agreed/strongly agreed that MAP was useful for managing their asthma.

Discussion   Ongoing access by participants supports the feasibility of using the system to deliver self-management for individuals who accept to start using the system. Given that logins decreased over time, this may reflect that once individuals learned about their medications and action plan they found less benefits of logging in to report symptoms and receive feedback that they became more familiar with over time.

Conclusion   The results of this process evaluation will be used to inform future changes to MAP including tailoring content and features over time to changes in individuals' knowledge, confidence and symptoms.

Abstract no. 540   Sharing medical images: challenging issues and lessons learned from pilot implementation at a tertiary university hospital in South Korea

Taeki Kim, *Seoul National University Bundang Hospital, Seongnam-si, South Korea*

Sooyoung Yoo, *Seoul National University Bundang Hospital, Seongnam-si, South Korea*

Hyun Young Baek, *Seoul National University Bundang Hospital, Seongnam-si, South Korea*

Eunyoung Heo, *Seoul National University Bundang Hospital, Seongnam-si, South Korea*

Jeong-Whun Kim, *Seoul National University Bundang Hospital, Seongnam-si, South Korea*

Hee Hwang, *Seoul National University Bundang Hospital, Seongnam-si, South Korea*

Introduction   This study implemented a pilot of a picture archiving and communication system based on international standards. The medical environment of a tertiary university hospital was used.

Methods   For exchange of medical images, an authentication process on a portal site was required before the use of the service. The range of images such as computed tomography (CT), magnetic resonance imaging (MRI) and sonography was set for 4 general hospitals, 20 hospitals and 60 clinics, and the system at each hospital was assessed to be archived in the picture archiving system. Both standardised and non-standardised images from each of the hospitals were analysed and used for the development of the medical image exchange system. The developed system has been evaluated through questionnaire survey of 48 patients and 17 medical staff who used the system from March 14 to 31, 2016.

Results   For the exchange of the image data in selected hospitals, an integrated system based on the Integrating the Healthcare Enterprise (IHE) Cross-Enterprise Document Sharing standards (XDS.b, XDS-I.b) and Digital Imaging and Communications in Medicine (DICOM) was established, which enabled the transfer and retrieval of the images. The ISO27001 standard was used for system security. The hospital administrators could transfer the images to the system only for the patients who had given their consents, and the administrators also had to be registered according to the standardised codes of the hospital. The physicians at the hospital could use the Patient Identifier Cross-Referencing to check the patients' previous images. They were able to download the images of the patients if necessary and upload them on the picture archiving and communication system (PACS) of the hospital to request for readings and use them for diagnosis. As a result of the survey conducted after using the system, 94% of respondents answered that the medical service was quick and accurate and 96% of them answered that it shortened the treatment period. 71% of the medical staff responded positively to the rapid diagnosis, 65% of them responded to the accuracy of the diagnosis and 100% of the medical staff responded that they were able to avoid the double check.

Discussion   This system had many benefits for patient diagnosis but there were some limitations in the procedural aspects. Some patients were capable of ordinary daily life but those who used the image exchange system mostly had multiple examinations previously most were emergency patients or patients of chronic illnesses with histories of diagnosis in primary, secondary and tertiary hospitals and were in the older age group who visited the hospitals with the help of their guardians.

The use of this service to transfer and retrieve patient images requires an informed consent, so it was necessary to enhance the convenience of doing this.

Conclusion   The system developed in this study could be used by all subjects, and patients could use their previous image data from anywhere in Korea to assist their diagnosis.

If medical record and image exchange becomes integrated in the future, it would help prevent and diagnose patients, while supporting the record management and eradication at times of national pandemic. This work was supported by the IT R&D program of Seoul National University Bundang Hospital.

Abstract no. 545   Semantic technologies for improved primary and secondary use of clinical data

Stefan Schulz, *Institute for Medical Informatics, Statistics and Documentation, Medical University of Graz, Graz, Austria*

Markus Kreuzthaler, *Institute for Medical Informatics, Statistics and Documentation, Medical University of Graz, Graz, Austria*

Diether Kramer, *CBmed - Center for Biomarker Research, Graz, Austria*

Werner Leodolter, *Steiermärkische Krankenanstalten, Graz, Austria*

Markus Pedevilla, *Steiermärkische Krankenanstalten, Graz, Austria*

Berthold Huppertz, *Steiermärkische Krankenanstalten, Graz, Austria*

Karine Sargsyan, *Biobank Graz – Medical University of Graz, Graz, Austria*

Dagmar Becker, *Biobank Graz, Graz, Austria*

Robert Fasching, *SAP, Walldorf, Germany*

Thomas Pieber, *CBmed, Graz, Austria*

Introduction   Heterogeneous content of electronic health records are increasingly considered a promising data treasure for primary and secondary use. The Austrian Center for Biomarker Research in Medicine investigates and implements methods for a seamless integration of clinical and research data.

Method   Structured and unstructured content is standardised and analysed, using natural language processing, terminology standards (e.g. Systematized Nomenclature of Medicine - Clinical Terms (SNOMED CT)), big data management and predictive content analytics. Semantically explicit data extracts are stored in a data warehouse based on SAP-HANA (Systems, Applications & Products [in Data Processing] High-performance ANalytic Appliance). Four different application scenarios are being investigated and implemented.

   (i) *Recruiting* facilitates the creation of patient cohorts and related data sets according to semantic filtering criteria, using graphical interfaces for querying and visualization.
  (ii) *Prediction* applies predictive analytics methods to semantically enriched patient profiles, in order to estimate the probability of future events, for example, hospital re-admissions.
 (iii) *QuickView* will provide an automatic summary of decision-relevant patient data, depending on the preferences of user groups and tasks.
 (iv) *Coding* will facilitate the assignment of administrative codes to care episodes, triggered by routine data.

Results   For scenario (i), a first retrieval and information extraction show case (defined by Biobank Graz) could be successfully integrated into SAP HANA, with structured data produced by an NLP pipeline analysing de-identified free-text discharge summaries. For (ii), prediction scenarios were defined and prototypically implemented using supervised and unsupervised machine learning. So far, prediction methods were tested on re-admission risk, risk of delirium in geriatric patients and probable comorbidity ICD (International Classification of Diseases) codes. For scenario (iii), interviews with physicians have been carried out, resulting in a first design and implementation mock-up for patient-based navigation and summarization. Finally, scenario (iv) has just started with collecting requirements.

Discussion   Early results in scenarios (i)–(iii) are promising, but numerous challenges will have to be addressed in this long-term project, such as heterogeneous data landscape, local medical terminology, semantic standards, diverse information needs, data quality and user acceptance and performance.

Conclusion   A high-performance data warehouse technology, together with semantic technologies, opens new frontiers for massive clinical data (re)use. However, all use cases need to be carefully evaluated so that they can be applied in a clinical setting.

Abstract no. 588   Automating clinical pathways using executable business process model and notation

Vishnu Vardhan Chandrabalan, *Manchester Royal Infirmary, Central Manchester Foundation Trust, Manchester, UK*

Mohan Shanmugham, *Manchester Royal Infirmary, Central Manchester Foundation Trust, Manchester, UK*

Introduction   Clinical pathways are similar to business process in many aspects. Well-structured clinical workflows improve the efficiency and safety of healthcare delivery. Business Process Model and Notation (BPMN) is an open standard for the graphical representation of business processes. Executable BPMN deployed using a 'process engine' allows for interactive and semi-automated processing of user tasks and decision-making.

Patients admitted with acute surgical conditions requiring emergency surgery follow predictable clinical pathways. Optimisation of these pathways not only improves patient experience and outcome but can also result in significant healthcare cost benefits. The aim of this abstract is to demonstrate a 'proof-of-concept' application of executable BPMN to an emergency surgery clinical pathway.

Methods   Healthcare tasks undertaken during the course of treatment of a patient with a surgical emergency were listed. These tasks were ordered in a time-based sequence and were categorised based on the healthcare professionals responsible for delivering these tasks. Exclusive and parallel gateways were determined using clinical decision nodes. These were used to create a BPMN model using an open-source modeller.

Task forms were designed to capture information that was relevant to the movement of the patient along the pathways. User roles were defined and tasks were assigned to users as would happen in a real-world clinical scenario. Automated notifications were used to inform users of pending tasks.

Results   The BPMN process model designed for an emergency surgery clinical pathway is shown below. HTML forms were designed for each task allowing capture of data relevant to decision making as well as to automate progression along the pathway by facilitating automated communication between team members involved in the patient's care.

Discussion   Executable BPMN can be applied easily to healthcare processes with effective integration into existing clinical workflows. Such integration not only allows for effective data capture, communication between teams and improved efficiency,
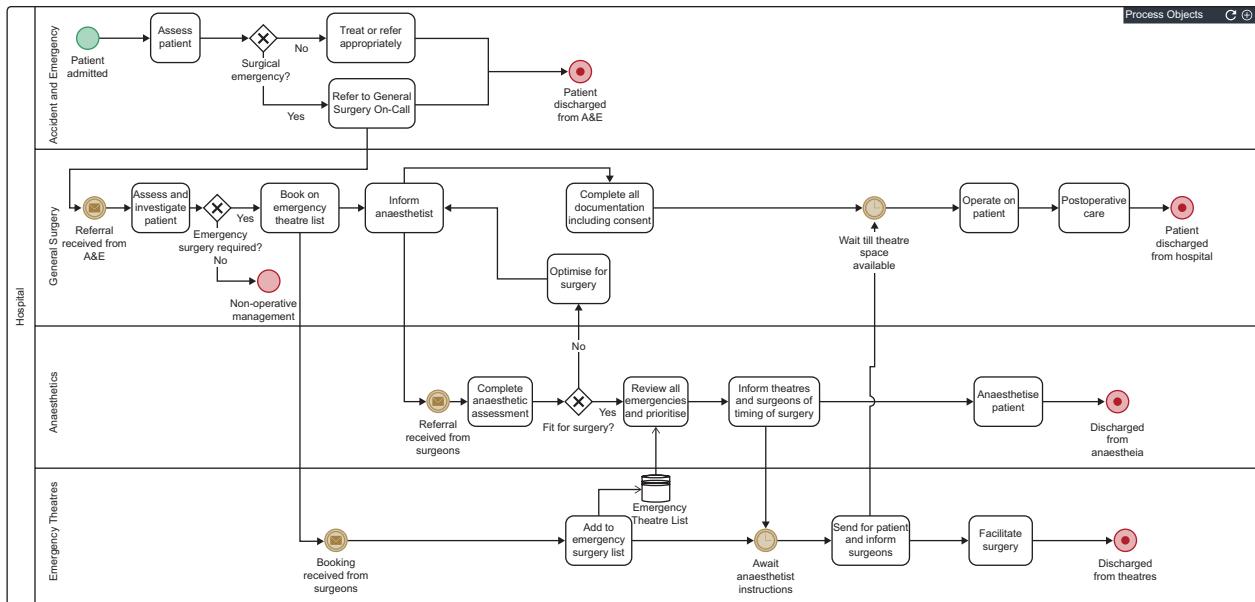
Figure 1 BPMN process model for emergency surgery pathway

but also allows for analysis of bottle-necks and organisational optimisation. Dashboard views of overall workload and activities in progress allow healthcare providers and managers to deploy resources in a more dynamic manner in real-time.

Conclusion   Executable BPMN with interactive forms run using a process engine can be easily adapted to routine healthcare processes to produce significant improvements in efficiency with minimal investment in IT resources.

Abstract no. 589   GlobalSurg: enabling global health research in surgery

Thomas Drake, *University of Edinburgh, Edinburgh, UK*

Riinu Ots, *University of Edinburgh, Edinburgh, UK*

Catherine Shaw, *Royal Infirmary of Edinburgh, Edinburgh, UK*

Ewen Harrison, *University of Edinburgh, Edinburgh, UK*

Introduction   The provision of safe surgery across the world has been recently identified as a key area that requires urgent attention to reduce health inequalities. There is a specific lack of data across the world on infection following surgery, which is a key driver of morbidity. This study aimed to describe the epidemiology of surgical site infection following gastrointestinal surgery across the world.

Methods   Data were collected by clinicians across the world, through a network of senior clinicians who were responsible for their country. Collected data were subsequently validated by investigators at each centre to ensure accuracy. Data entry was facilitated via a web-based server platform (REDCap). An analysis workflow was then enabled using an API, permitting real-time data visualisation and study management. Embedded sub-studies also tested the feasibility of asynchronous mobile data upload in low-middle income countries (LMICs).

Results   In total, 376 hospitals across 65 countries (30 high, 18 Middle and 17 low income countries) collected data on 15,936 patients. Data collection via a mobile platform with asynchronous upload reduced the time required to enter data, the resources required and proved feasible in an LMIC setting.

Discussion   This study demonstrated collecting patient level data on an international level is feasible, even in low resource settings. Use of mobile platforms with asynchronous upload facilities reduces the workload and resources required to achieve this. Future research should focus on deploying new technologies to LMICs, including machine learning and feedback to clinicians for quality improvement.

Abstract no. 590   A precision medicine approach to the prediction of kidney stones formation for an at-risk population of individuals admitted to the Emergency Room

Zhaoyi Chen, *University of Florida, Gainesville, FL, USA*

Victoria Bird, *University of Florida, Gainesville, FL, USA*

Mattia Prosperi, *University of Florida, Gainesville, FL, USA*

Introduction   Kidney stones are detrimental and painful and carry a high cost to the health care system. Large-scale, long-term electronic health record (EHR) data bases of patients presenting to the Emergency Room (ER) can be exploited to identify risk patterns of developing kidney stone formation (KSF). Our objective is to study a large 12-year ER data base and infer a statistical model for the risk of KSF. The ultimate goal is to make a personalised predictive tool for the population presenting to ERs with urinary symptoms as well as the general population.

Methods   Data from the MIMIC-III EHR data base included 38,158 adult subjects who visited the ER between 2000 and 2012. We selected patients diagnosed with KSF, n = 216 compared to those who had other urinary conditions, n = 14,358 and controls, n = 23,584 (all other conditions). Statistical analysis was conducted on demographic characteristics, Charlson Comorbidity Index (CCI), other disease diagnoses, charted events and lab tests. ICD diagnoses and lab/chart events were identified using univariate analysis (Bonferroni-adjusted). Multivariable (with forward selection) logistic regression models were fitted, and discriminative ability of models were tested using under the receiver operating characteristic (ROC) analysis with 30% holdout test. All statistical analysis was conducted using SAS 9.4.

Results   Differences in age and gender between KSF and other two groups were observed. The highest rates of KSF were seen in the age group 71–80 for both males (29.59%) and females (22.8%). KSF tends to be associated with the overweight category for males and obese for females. Among CCI categories, male and female KSF had significantly lower CCI of 0–1 compared to other groups (p < 0.0001). Among all ICD diagnoses that significantly differently distributed between KSF and other two groups, essential hypertension (45.2%), disorders of fluid, electrolyte and acid-base balance (44%), septicaemia (41.7%), atelectasis and acute respiratory distress syndrome (30.1%), heart failure (25%) has the highest frequencies in KSF. The area under the ROC obtained on holdout testing data was 87% for model with lab/chart event, 78% for model with ICD diagnosis and 65% for model with the CCI. The sensitivity was about 70% and 80% at 75% specificity for model with ICD diagnosis and model with lab/charted events respectively.

Discussion   This data showed different presentation patterns in terms of age, gender, body mass index and CCI. Some unexpected diagnoses and lab/chart events were identified to be associated with KSF, and future studies could focus on address these associations and to discover potential underlying mechanism. In terms of predictive ability, multivariable models can discriminate well KSF, but the low sensitivity and specificity warrant investigation of other potential markers, especially for early detection.

Conclusion   We examined the risk patterns of developing KSF in an ER/ICU setting, and several ICD diagnoses, lab tests and charted events potentially associated with KSF were identified. Further model optimization is needed to improve the predictive ability and generalizability of our models.

Abstract no. 591   GIS mapping of Dengue incidence in Punjab, Pakistan

Suleman Atique, *Taipei Medical University, Taipei, Taiwan*

Chien-Yen Hsu, *National Taipei University of Nursing and Health Sciences, Taipei, Taiwan*

Syed-Abdul Shabbir, *Graduate Institute of Biomedical Informatics, Taipei Medical University, Taipei, Taiwan*

Introduction   Dengue has become a great threat to public health in tropical and sub-tropical regions, leading to 2.5 billion, almost half of world's population, people at risk of dengue fever on the globe as estimated by World Health Organization and case fatality rates vary between 0.5 and 3.5% in Asian countries. There were approximately 8.3 million cases of dengue fever in 1990 and 58.3 million apparent cases in 2013 and more than two billion people are at risk of being infected with dengue virus leading to about 55,000 people requiring hospitalization. Majority of these who require hospitalization are children.

Methods   Dengue case data from 2006-2014 have been collected from provincial health department of Punjab province. Data for the estimated population of each district have been downloaded from the website of 'Bureau of Statistics, Punjab'. Arc GIS

10.2 ArcGIS® software by Esri is used to geographically visualise the incidence of dengue fever according to each year. It visualises the dengue incidence rate in each of the 36 districts in Punjab.

Results   The results show that in 2006 dengue was present only in Lahore district. The trend was pretty similar till 2010, and it did not spread much to other areas of Punjab. However, in 2011, there was a huge outbreak of dengue fever across the whole province with more than 20 thousand cases and 350 deaths.

Discussion   Dengue is spreading its territory because of viable conditions. Maps indicate its potential and shift from southern region towards to northern region of Punjab.

Conclusion   It can be concluded from the study that dengue is rapidly expanding its territory across the country and proper vector control measures should be implemented to stop its spread. There is need to spread awareness among the public and integrate vector control strategies to control dengue expansion.

Abstract no. 605   Risk factors for incident heart failure in a population-based cohort using linked electronic health records (CALIBER)

Alicia Uijl, *Julius Center, University Medical Center Utrecht, Utrecht, The Netherlands*

Stefan Koudstaal, *Farr Institute of Health Informatics Research, Institute of Health Informatics, University College London, London, UK*

Kenan Direk, *Farr Institute of Health Informatics Research, Institute of Health Informatics, University College London, London, UK*

Spiros Denaxas, *Department of Cardiology, Division Heart & Lungs, University Medical Center Utrecht, Utrecht, The Netherlands*

Rolf Groenwold, *Farr Institute of Health Informatics Research, Institute of Health Informatics, University College London, London, UK*

Arno Hoes, *Institute of Cardiovascular Science, Faculty of Population Health Sciences, University College London, London, UK*

Folkert Asselbergs, *Farr Institute of Health Informatics Research, Institute of Health Informatics, University College London, London, UK*

Background   Heart failure is one of the leading causes of mortality and morbidity among patients aged 55 years and older. Several risk factors for incident heart failure have been identified; however, we hypothesise there might be risk factors yet to be described due to limited information on heart failure risk factors and comorbidities in previous studies. Linked electronic health records (EHR) that capture clinical data across healthcare settings may provide the opportunity to discover and examine previously unknown risk factors across different sub-groups from the general population at high-risk for the development of heart failure.

Aim   To characterise known and unknown risk factors for incident heart failure in the general population 55 years and older and to quantify their relative contribution.

Methods   We used linked EHR data from 2000 to 2010 as part of the UK-based CALIBER resource, in which clinical data from primary care (Clinical Practice Research Datalink), hospital admissions (Hospital Episode Statistics) and mortality (Office for National Statistics) are linked for research purposes. We assembled a cohort of individuals 55 years or older and free of heart failure at baseline. Potential risk factors included: ethnicity, social deprivation (index of multiple deprivation), body mass index, physical activity, smoking, alcohol consumption, blood pressure, lipid levels and comorbidities such as hypertension and diabetes mellitus. Informed by the literature, we selected several other potential risk factors that may have a role in modifying heart failure outcomes. The primary endpoint was the first record of heart failure from primary care or secondary care. Events in primary care were defined by a diagnosis of heart failure or notification of left ventricular dysfunction on echocardiogram, and in secondary care by a diagnosis of heart failure. We have characterised the diverse case-mix of incident heart failure by gender and age. Multivariable Cox regression analysis will be used to estimate hazard ratios for associations between risk factors and incident heart failure, separately for men and women and by age category: 55–64 years, 65–74 years, 75–84 years and >85 years.

Results   We identified 859,481 individuals, of whom 397,454 (46.2%) were male. Median age at cohort enrolment was 61.5 years. During a median follow-up of 5.7 years, a total of 48,144 incident heart failure events were recorded. Of the 23,380 (48.6%) males with incident heart failure events, 23.2% were between 55 and 64 years, 34.5% between 65 and 74 years, 33.4% between 75 and 84 years and 8.8% were 85 years or older. Of the 24,764 (51.4%) women with incident heart failure events, 11.7% were between 55 and 64 years, 26.8% between 76 and 74 years, 41.5% between 75 and 84 years and 20.1% were 85 years or older. Main findings will be presented at the conference.

Abstract no. 633   ePrescribing – how does it affect reported medication errors?

Claire Thompson, *University of Leeds, Leeds, UK*

Ruth Evans, *University of Leeds, Leeds, UK*

Introduction   The adoption of information technology to support the delivery of safe, efficient healthcare is increasing, replacing paper health records with electronic systems. Within Leeds Teaching Hospitals NHS Trust (LTHT), ePrescribing is being implemented with the aim patient safety, clinical effectiveness and operational productivity. LTHT is an acute hospital trust with 2500 inpatient beds and is a regional and national centre for specialist treatment. The implementation of ePrescribing commenced in September 2015. These mixed methods pilot study evaluated the effect of the implementation of ePrescribing on reported medication errors by early system adopters. The analysis of medication error reports in these areas will refine a method that can be used as the system is rolled out further in order to determine conclusions across the entire trust.

Methods   Medication error reports were analysed from the seven wards live with ePrescribing and included a mix of specialties across two different sites. Anonymised error report data was extracted from the LTHT Datix Incident Reporting system for a period of 42 ward-months pre-implementation and 38 ward-months post-implementation. Error reports were categorised by medication error type and stage of medication error occurrence using a method developed following a literature review, and from World Health Organisation guidance on medicines errors. Semi-structured interviews were conducted with four-ward management staff in areas where ePrescribing has been implemented.

Results   A total of 90 medication errors were analysed. 64 were recorded pre-system implementation, and 26 errors were recorded post-implementation. Pre-implementation, the mean error rate was 1.5 errors per ward month. Post-implementation, this was reduced to 0.68 errors per ward month. The interviews were transcribed and analysed using a semantic analysis method.

Discussion   The categorisation of each medication error allows for in-depth analysis, in order to realise trends throughout the implementation process. For instance, during the analysis of medication errors, it was apparent that the medication error type 'drug or dosage omission', which was the highest occurring medication error pre-implementation, was non-existent post-implementation. During interviews with ward staff, it became apparent that the meaning of a drug or dosage omission is not the same pre-implementation as it is post-implementation. Pre-implementation, a blank square on a medication chart, would be reported as a drug omission during regular pharmacy reviews, meaning that the dose had been missed. Post-implementation, blank squares cannot exist as each administration, which has not had an action against it, will present to every user, requiring the dose to be marked as missed or withheld if not given.

Conclusions   This research concluded that the implementation of ePrescribing could reduce medication errors in some areas but that to demonstrate this more clearly, further data on medication errors should be analysed as the ePrescribing system is rolled out across the entire Trust, to give a wider view on the impact and over a longer period of time. It is recognised that the classification of medication errors is vital in order to support further research, as this will enable better understanding of the effect of ePrescribing on medication errors.

Abstract no. 651   Using the Nextflow framework for reproducible in-silico omics analyses across clusters and clouds

Maria Chatzou, *Center for Genomic Regulation, Barcelona, Spain*

Paolo Di Tommaso, *Center for Genomic Regulation, Barcelona, Spain*

Evan Floden, *Center for Genomic Regulation, Barcelona, Spain*

Cedric Notredame, *Center for Genomic Regulation, Barcelona, Spain*

Introduction   Reproducibility has become one of biology's most pressing issues. This impasse has been fuelled by the combined reliance on increasingly complex data analysis methods and the exponential growth of biological data sets. When considering the installation, deployment and maintenance of bioinformatic pipelines, an even more challenging picture emerges due to the lack of community standards. The effect of limited standards on reproducibility is amplified by the very diverse range of computational platforms and configurations on which these applications are expected to be applied (workstations, clusters, high-performance computing, clouds, etc.). With no established standard at any level, diversity cannot be taken for granted.

Methods   Nextflow is a pipeline orchestration tool that has been designed to ease deployment and guarantee reproducibility across platforms. It is a computational environment that provides a domain specific language to simplify the writing of complex distributed computational workflows in a portable and replicable manner. It allows the seamless parallelization and deployment of any existing application with minimal development and maintenance overhead, irrespective of the original programming language. The built-in support for container technologies such as Docker and Shifter, along with the native integration with the Git tool and popular code-sharing platforms like GitHub, make it possible to precisely prototype self-contained computational workflows, maintain all variations over time and rapidly reproduce any former configuration one may need to re-use. These capabilities guarantee consistent results over time and across different computing platforms.

Results   Using a simple RNA-Seq-based analysis, we show how two seemingly irreproducible analyses can be made stable across platforms when ported into Nextflow. Applying the transcript quantification tool Kallisto and the companion differential expression package Sleuth, we highlight how the same pipeline produces different results on Mac OSX and Linux. With a well-studied data set, 67 genes are only reported as significant on the Mac and 72 only reported in the Linux analysis out of a total of 6,138 differentially expressed genes. When the same pipeline was factored into Nextflow with Docker, the results were identical across platforms.

Nextflow was first introduced at BOSC 2015. Over the last year, it has attracted an increasing amount of interest within the bioinformaticians community, including numerous users from leading institutions such as the Joint Genome Institute, the Pasteur Institute, the Sanger Institute, the International Agency for Research on Cancer and the Weill Cornell Medical College among others. An exhaustive list of third party curated pipelines is available from https://github.com/nextflow-io/awesome-nextflow

---

Abstract no. 677   Developing a set of administrative case definitions for identifying sleep disorders in ICD-coded data

Rachel Jolley, *Department of Community Health Sciences, University of Calgary, Calgary, AB, Canada*

Jane Liang, *Department of Community Health Sciences, University of Calgary, Calgary, AB, Canada*

Guanmin Chen, *Department of Community Health Sciences, University of Calgary, Calgary, AB, Canada*

Danielle Southern, *University of Calgary, Calgary, AB, Canada*

Hude Quan, *Department of Community Health Sciences, University of Calgary, Calgary, AB, Canada*

Introduction   Sleep disorders are common and have been shown to have large impacts on health – associated with increased morbidity and mortality. There are no guidelines or evidence to inform the identification of a case definition of sleep disorders from ICD coded administrative data. The objectives of this study are to develop and validate an administrative data coded algorithm to define sleep disorders including narcolepsy, insomnia and obstructive sleep apnoea in Alberta, Canada in a sleep-disorder cohort.

Methods   A cohort of previously reviewed adult patient records from a sleep clinic in Calgary, Alberta, Canada, between 1 January 2009 and 31 December 2011 was used as the reference standard medical record review. We developed a general ICD-10 case definition of sleep disorders that included conditions of narcolepsy, obstructive sleep apnoea and restless leg syndrome using administrative data including: 1) physician claims data, 2) in-patient visit data (discharge abstract data) and 3) emergency/ambulatory care data (Ambulatory Care Classification System (ACCS), National Ambulatory Care Reporting System (NACRS)). We linked the medical record review data and administrative data through a unique personal health identification number to examine validity. Different case definitions were tested with estimates of sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) calculated.

Results   From a total of 1209 patients, 1085 (89.7%) were classified has having a sleep disorder and 124 (10.3%) were classified as not having a sleep disorder. The most optimal definitions within each data set included physician claims data definition: a) within three years of the clinic visit with 1 positive ICD-10 code showed sensitivity = 79.8%, specificity = 11.7%, PPV = 90.1% and NPV = 11.7 and b) 1 year prior to sleep clinic visit +3 years after visit with 1 positive ICD-10 code showed sensitivity = 86.8%, specificity = 16.9%, PPV = 90.1% and NPV = 12.8%. Emergency/ambulatory care data (ACCS and NACRS) data definition: a) within three years of the clinic visit with 1 positive ICD-10 code showed sensitivity = 96.4%, specificity = 16.1%, PPV = 90.7% and NPV = 33.9% versus definition, b)1 year prior to sleep clinic visit +3 years after visit with 1 positive ICD-10 code showed sensitivity = 99.4%, specificity = 8.9%, PPV = 90.5% and NPV = 61.1%. The in-patient data yielded poor results in all year and positive code combinations.

Discussion   In the cases of physician claims and emergency/ambulatory care data, the sensitivity of the ICD-10 coded algorithm was relatively high, as was the PPV; however, the specificity and NPV were low irrespective of the data source or definition. The sensitivity increased in both data sets of the ICD-10 codes but this was dependent on the timing of the definition, specifically including the year prior to the sleep clinic visit, increased the sensitivity and PPV. This may be due to the way in which sleep disorders are diagnosed in terms of GP or out-patient visits prior to the referral to a sleep clinic for further diagnosis.

Conclusions   This study demonstrates that sleep disorders in administrative data are being captured; however, the specificity of the definition is poor. In order to optimise the definitions, the data sources should be linked and those with both high sensitivity and specificity should be preferentially considered if used for surveillance purposes.