Scottish Informatics Programme

Supported by wellcome trust

Exploiting Existing Data for Health Research

University of St Andrews

International Conference
28th – 30th August 2013

Conference Programme and Abstracts

Programme: Wednesday 28th August

11.00- 12.45 pm: REGISTRATION and BUFFET LUNCH
Venue: Lower and Upper college Halls

12.45-1.00 pm: WELCOME ADDRESS, Chris Dibben (University of St Andrews and Director of the Longitudinal Studies Centre – Scotland and the Administrative Data Liaison Service)
Venue: The Younger Hall

1.00-1.55 pm: KEYNOTE: "Big Data meets Healthcare: The case for comparability and consistency" Professor Christopher G. Chute, Professor of Medical Informatics, Mayo Clinic and Chaired by Dr Chris Dibben
Venue: The Younger Hall

2.00 - 3.30 pm: CONFERENCE SESSION 1

a) Vulnerable children and young adults ***Venue*: *School I***
b) Routine data for neonatal research ***Venue*: *School II***
c) Examining patterns of service utilisation with routine or linked data *Venue*: ***School III***
d) Mental health ***Venue*: *School V***
e) Methodological Advances and New Initiatives in Data Linkage ***Venue*: *Irvine Lecture Theatre***
f) Data transparency, access and public engagement in routine health data research ***Venue: Forbes Lab***
g) Routine data for RCT 1 ***Venue*: *Room 31***

3.30-3.50 pm: TEA/COFFEE *Lower and Upper college Halls*

4.00-4.55 pm: PANEL DISCUSSION PLENARY: "Future directions and priorities in UK e-Health research" Contributors: Professor Ronan Lyons, Professor Iain Buchan, Professor Harry Hemingway, Professor Andrew Morris and Dr. Chris Dibben. Chaired by Dr. Tim Hubbard
Venue: The Younger Hall

5.00-6.30 pm: CONFERENCE SESSION 2

a) Social and demographic dimensions of health ***Venue*: *School I***
b) Childbirth ***Venue*: *School II***
c) Co-and multi-morbidity ***Venue*: *School III***
d) Cancer research using data linkage 1 ***Venue*: School V**
e) Analysis of routine administrative and linked data *Venue*: ***Irvine Lecture Theatre***
f) Exploring consent and engagement amongst participants ***Venue*: *Forbes Lab***
g) Routine data for RCT 2 ***Venue*: *Room 31***

8.00 pm: SOCIAL EVENT: "'We'll tak a cup of kindness yet', The flavour of Scotch Whisky", Followed by a whisky tasting session, Led by Dr. David Wishart
Venue: The Ballroom, The Old Course Hotel

Programme: Thursday 29th August

9.00-10:30 am: CONFERENCE SESSION 3

a) Facilitating the use of routine health data ***Venue: Irvine Lecture Theatre***
b) Access to healthcare and the spatial dimensions of health ***Venue*****:** ***School III***
c) Examples of E-Health research 1 ***Venue*****:** ***Forbes Lab***
d) Care for older people ***Venue*****:** ***Room 31***
e) Linkages to enhance existing data 1 ***Venue*****:** ***School I***
f) International perspectives on data governance ***Venue*****:** ***School II***
g) Mental health among youth ***Venue: School V***

10:30-10.55am: TEA/COFFEE *Venue: Lower and Upper college Halls*

11.00am – 12.00 noon: CONFERENCE SESSION 4:

a) Alcohol, smoking and drug misuse 1 ***Venue: School I***
b) Child development and the early years ***Venue: Irvine Lecture Theatre***
c) Data linkage in environmental health studies ***Venue: School V***
d) Missing data in longitudinal e-Health databases ***Venue: School II***
e) Information Governance and NHS data: Navigating Legal Constraints in research ***Venue: School III***
f) Linked data for hard to reach groups ***Venue: Forbes Lab***

12.00 – 12.55 pm: LUNCH and POSTER PRESENTATIONS

Venue: Lower and Upper college Halls (Posters displayed in upper college hall)

1.00 -2.30 pm: CONFERENCE SESSION 5

a) Obesity and BMI ***Venue: School V***
b) Alcohol, smoking and drug misuse 2 ***Venue: Room 31***
c) Data linkage and access platforms ***Venue: School III***
d) National Cardiovascular Registries for Monitoring Quality of Care ***Venue: Irvine Lecture Theatre***
e) Special Session: REporting of studies Conducted using Observational Routinely-collected Data RECORD initiative ***Venue: School I***
f) Socio-economic inequalities in health ***Venue: Forbes Lab***
g) The value of data linkage: international perspectives ***Venue: School II***

2.30-2.55 pm: TEA/COFFEE *Venue: Lower and Upper college Halls*

3.00-4:30 pm: CONFERENCE SESSION 6

a) Interventions and health policy ***Venue: Forbes Lab***
b) Kidney illnesses ***Venue: Room 31***
c) Linking within families ***Venue: Irvine Lecture Theatre***
d) Examples of E-Health research 2 ***Venue: School I***
e) Using linked data for validation 1 ***Venue: School V***
f) Linkage methods ***Venue: School II***
g) Data management and meta-data: systems and platforms ***Venue: School III***

4:40-5:40 pm: KEYNOTE: "Reboot 2020" Sander Duivestein, Institute for the Analysis of New Technology followed by a response from Dr. Mark Elliot, University of Manchester**.** Chaired by Dr. Colin McCowan
Venue: The Younger Hall

7.00 pm: CONFERENCE DINNER and CEILIDH
Venue: Hall of Champions, Old Course Hotel, St Andrews

Programme: Friday 30th August

9.00-10:30 am: CONFERENCE SESSION 7

a) Data linkage approaches and evaluating linkage quality ***Venue: School III***
b) Integrating and curating heterogeneous clinical data, and transforming them into research-ready variables ***Venue: Forbes Lab***
c) Special Session Randomised trials utilising electronic health records ***Venue: School II***
d) Health-care costs and provision ***Venue: School V***
e) Enhancing routine health data ***Venue: School I***
f) Challenges and solutions to the complexity of data linkage ***Venue: Irvine Lecture Theatre***

10.30-10.55: TEA/COFFEE *Venue: Lower and Upper college Halls*

11.00-12.00 noon: KEYNOTE: "Reusing historical data: the Scottish Mental Surveys of 1932 and 1947" Professor Ian Deary, University of Edinburgh and Chaired by Professor Frank Sullivan
Venue: The Younger Hall

12.00-12.55: LUNCH
Venue: Lower and Upper college Halls

1.00-2:30 pm: CONFERENCE SESSION 8:

a) Special session: The role of computer science in E-health research ***Venue: Forbes Lab***
b) Child Health ***Venue: School I***
c) Governance: Ethics and training strategies ***Venue: School II***
d) Linkages to enhance existing data 2 ***Venue: School III***
e) Cancer research using data linkage 2 ***Venue: School V***

2:30-3.00pm: TEA/COFFEE *Venue: Lower and Upper college Halls*

Depart

Session Details and Paper Titles

Session 1: 2.00 - 3.30 pm, Wednesday 28th August						
Session 1a Vulnerable children and young adults *(School I)* **Chair: Zhiqiang Feng** Linking data to build an evidence base- The Developmental Pathways Project *De Klerk* Comparison of incidence trends in victimisation-related injury admission in children and adolescents, between England and Scotland. *Gonzalez-Izquierdo* Entering out-of-home care during childhood: Cumulative incidence study in two developed countries *O'Donnell* Do NEET	**Session 1b Routine data for neonatal research *(School II)*** **Chair: Nirupa Dattani** Establishing a National Neonatal Research Database from operational NHS electronic records *Statnikov* The UK Neonatal Collaborative - Necrotising Enterocolitis Study: a prospective population-based study using the National Neonatal Research Database *Battersby* Risk factors for infant deaths among singleton babies born at term in England, 2005-07 *Dattani* Creating the Scottish Congenital Anomalies	**Session 1c Examining patterns of service utilisation with routine or linked data *(School III)*** **Chair: Frank Sullivan** Use of multiple electronic data capturing systems to monitor patients at a major trauma centre *Nisar* Antidepressant prescriptions and subsequent health service utilization in Austria: A record linkage study in a country with a fragmented payment system and only partially available unique patient identifiers *Katschnig* The effect of primary care psychologist treatment on health	**Session 1d Mental health *(School V)*** **Chair: Colin McCowan** Are psychotic experiences pathological? Comparison between information on self-reported and observer-rated psychotic experiences in the ALSPAC birth cohort study and clinical information in linked primary care records *Davies* Psychotropic medication utilisation in care home residents age 65 or older compared with the equivalent general population in Scotland *McTaggart* Can the use of routine data enhance collection of the primary outcome in the SHIFT Trial	**Session 1e Methodological Advances and New Initiatives in Data Linkage *(Irvine Lecture Theatre)*** **Chair: James Boyd** Beyond “Big Data”: assessing the quality and utility of administrative data for research use *Smith* Improving linked data quality, research outcomes and reducing costs using graph theory and graph databases *Farrow* Sampling based clerical assessment *Guiver* ESRC Welsh Government Data Linking Demonstration Projects 2012-13: Methodological Challenges and Lessons Learned *Lowe*	**Session 1f Data transparency, access and public engagement in routine health data research *(Forbes Lab)*** **Chair: Mhairi Aitken** Implementing safe effective proportionate governance in NHS National Services Scotland *Murray* Control and Trust as the Foundations of Public Acceptability: Reflections from the Public Engagement workstream of SHIP *Aitken* Using Electronic Health Records for research purposes: key findings from a survey on patient and public attitudes *Papoutsi* The Freedom of Information Act (2000) in healthcare research:	**Session 1g Routine data for RCT 1 *(Room 31)*** **Chair: Ian Ford** The Candida in Pregnancy Study (CiPS): a randomised controlled trial [ACTRN12610000607077] *Roberts* Navigating the challenges of record linkage in a multi-centre research study across two health boards *Strachan* Using linked healthcare data to create a randomised controlled trial: The RAPiD Trial (Reducing Antibiotic Prescribing in Dentistry) *Elders* The use of routinely collected health data in clinical trials: an example from the SELENIB bladder

experiences have adverse impacts on health? Evidence from Scotland. *Feng*	Linked Dataset: A New Methodology *Nolan*	care utilization in general practice. A longitudinal data linkage study. *Prins* Body mass index and hospital admissions in UK women: a prospective cohort study *Reeves*	*Wright-Hughes* Results of depression screening in large community based population *Purves*		A Systematic Review *Fowler*	cancer trial *Evison*

Session 2: 5.00-6.30 pm, Wednesday 28th August

Session 2a Social and demographic dimensions of health ***(School I)***	**Session 2b Childbirth** ***(School II)***	**Session 2c Co-and multi-morbidity** ***(School III)***	**Session 2d Cancer research using data linkage 1** **(School V)**	**Session 2e Analysis of routine administrative and linked data** ***(Irvine Lecture Theatre)***	**Session 2f Exploring consent and engagement amongst participants** ***(Forbes Lab)***	**Session 2g Routine data for RCT 2** ***(Room 31)***
Chair: Lee Williamson	**Chair: Alison Macfarlane**	**Chair: Marion Bennie**	**Chair: Corri Black**	**Chair: Anoop Shah**	**Chair: Claudia Pagliari**	**Chair: Colin McCowan**
Understanding the impact of fertility history on outcomes in mid-life in Scotland, a longitudinal approach using the Scottish Longitudinal Study (SLS) *Williamson* Social Housing and	Using linked data to analyse rates of intervention in childbirth in England by mothers' countries of birth *Macfarlane* Caesarean section rates? Using routinely collected data to examine inter-hospital variation. *Roberts* Inherited Risk of Pre-	The benefits of using linked primary and secondary care data in developing a population based co-morbidity score *Crooks* Charlson scores derived from administrative data and case-note review compared favourably in a population-based cohort *Johnston*	Burn injury and cancer risk: A record-linkage study using data from Western Australia and Scotland. *Duke* Mobile phone use and risk of brain neoplasms and other cancers: prospective study. *Benson* Agricultural land usage	An Automated Method for Longitudinally Validating the Presence of Individuals in a Data Set *Thayer* Applying missing data methods to routine data: A prospective, population-based register of people with diabetes. *Read*	Consent to Data Linkage in the Context of the Avon Longitudinal Study of Parents And Children (ALSPAC): A Qualitative Study *Kennedy* Patterns in consent to study enrolment and linkage to health and	Linkage of routine data to generalise results from randomised controlled trials *Harron* Using routinely collected data to enhance long term follow up data: an example from the Building Blocks trial. *Cannings-John*

Health in Manitoba: A first look *Smith* Ethnic differences in upper gastrointestinal disease in Scotland *Ward* The Effect of Ethnicity on Outcomes Following Emergency Surgery: A national cohort study *Evison*	eclampsia: using two approaches for analysis *Bhattacharya* Recurrence Risk of Obstetric Anal Sphincter Injury (OASI) During Childbirth in New South Wales, Australia. *Ampt*	Multimorbidity: impact on health systems and their quality of care *Yu* Using hospital data to identify comorbidity and multimorbidity in Australia *Lujic*	and primary bone cancer: is there a link? Small-area analyses of osteosarcoma and Ewing sarcoma diagnosed in 0-49 year olds in Great Britain, 1985-2009 *Blakey* Policy for home or hospice as the preferred place of death from cancer: Scottish Health and Ethnicity Linkage Study shows challenges across all ethnic groups in Scotland *Sharpe*	Extracting information for research from free text in electronic health records *Shah* Methods for improving the estimation of multilevel survival models for large linked datasets *Stewart*	administrative records; evidence from the PEARL consent campaign *Davies* Is the use of linked routinely acquired NHS data for pharmacovigilance in children acceptable to parents and young people? *Scobie-Scott* Involving consumers in the work of a data linkage research unit *Jones*	Use of electronic health records to implement a cluster randomised trial in primary care. *Dregan* The importance of core outcome sets for practitioners, patients, policy makers and researchers *Williamson*

Session 3: 9.00-10:30 am, Thursday 29th August						
Session 3a Facilitating the use of routine health data *(Irvine Lecture Theatre)* **Chair: Tito Castillo** Using G-Cloud services to build a secure record linkage service *Castillo*	**Session 3b Access to healthcare and the spatial dimensions of health *(School III)*** **Chair: Chris Dibben** Exploring Social Segregation through urban form indicators and existing health data *Pasino*	**Session 3c Examples of E-Health research 1 *(Forbes Lab)*** **Chair: Mome Mukherjee** Estimation of familial effects on hospitalisation for common childhood infections *de Klerk*	**Session 3d Care for older people *(Room 31)*** **Chair: Iain Atherton** Use of linked primary and secondary care data to improve incidence estimates of community-acquired pneumonia in older adults in England *Millett*	**Session 3e Linkages to enhance existing data 1 *(School I)*** **Chair: Brad Kirby** Enhancing the intensively phenotyped and genotyped Generation Scotland Scottish Family Health Study cohort through record linkage.	**Session 3f International perspectives on data governance *(School II)*** **Chair: Nancy Meagher** The consultation on the establishment of a national health sector privacy advisory committee (NPAC) on	**Session 3g Mental health among youth *(School V)*** **Chair: Andy Boyd** Socioeconomic disadvantage, parental mental illness and deliberate self-harm in adolescents: a nested case-control study using record linkage *Hu*

Overcoming complexity and tedium: an object-oriented programming framework for linked data researchers. *Churches* An Introduction to the SAIL Databank *Jones* Role of the Research Co-ordinator *Nogueira*	Exploring the multidimensional influence of access to care on potentially preventable hospitalisations *Falster* Linking spatial accessibility of GP services to diabetes treatment *Atkinson* Online interactive atlas on chronic diseases and mental disorders *Vanasse*	Primary care data linkage to investigate risk factors associated with emergency hospital admission for Chronic Obstructive Pulmonary Disease *Hunter* Making sense of MS using linked data *Jones* An investigation into the use of aspirin and newer antiplatelets medications in Scotland following acute myocardial infarction *McTaggart*	The impact of first and second eye cataract surgery on injurious falls that require hospitalisation: a whole population study *Meuleners* Pathways in aged care: what we can learn from record linkage *Dickinson* Different effects of age, adiposity and physical activity on the risk of ankle, wrist and hip fractures in postmenopausal women: UK cohort linked to hospital admissions databases *Armstrong*	*Linksted* Navigating record linkage in Scotland and England and Wales: reviving the 6-Day Sample study *Brett* Data linkage for pharmacovigilance using routine electronic health records *Kirby* Use of record linkage for large-scale, epidemiological research: experience of UK Biobank *Sudlow*	the use of Scottish health records for research, statistical and related purposes. *Ruddy* Research Access to Health Administrative Data in Canada: Timelines, Processes, Successes and Bottlenecks *Meagher* Cross-jurisdictional Data Linkage: Lessons from Australia *Smith* Cross-country sharing of administrative health data: from aspirational to possible *Jorm*	Mental Health Service Utilization Among High Risk Youth *Blackadar* The likelihood of a child developing autism spectrum disorder, intellectual disability or both is related to a mother's mental health status in the years before the birth *Fairthorne* Prevalence of children's mental health disorders in survey data compared to population data: a comparison of two prospective cohorts *Wong*

Session 4: 11am – 12 noon, Thursday 29th August						
Session 4a Alcohol, smoking and drug misuse 1 ***(School I)*** **Chair: Mark McGilchrist** International	**Session 4b Child development and the early years** ***(Irvine Lecture Theatre)*** **Chair: Peter Donnan** Using routine data to	**Session 4c Data linkage in environmental health studies** ***(School V)*** **Chair: Tom Clemens** Linkage of Weather,	**Session 4d Missing data in longitudinal e-Health databases** ***(School II)*** **Chair: Irene Petersen** Missing data and multiple imputation in	**Session 4e Information Governance and NHS data: Navigating Legal Constraints in research** ***(School III)*** **Chair: Peter Coveney**	**Session 4f Linked data for hard to reach groups** ***(Forbes Lab)*** **Chair: Colin Fischbacher** Atrial fibrillation	**Free slot**

comparison of drugs users' cause-specific mortality needs to take account of epoch, demography, and injecting *Bird* Smoking, Surgery, and Venous Thromboembolism Risk in Women: UK cohort linked to hospital admissions databases *Green* Change in alcohol outlet density and alcohol-related harm to population health (CHALICE) *Fry*	identify developmental problems: the Childhood Information for Learning and Development project *Thompson* Prediction of initiation and cessation of breast feeding from late pregnancy to 16 weeks: The Feeding Your Baby (FYB) cohort study *Donnan* Using linked data sets to inform the early years agenda: infant feeding and child health in Scotland *Ajetunmobi*	Climate, and the Environment Data with Human Health and Wellbeing: MED-MI *Osborne* Address Cleaning and Temporal Linkage in Environmental Health Studies *Garwood* The association between ambient modelled air pollution and birth outcomes in Scotland *Clemens*	electronic health records *Bartlett* Longitudinal electronic health records and multiple imputation of missing data *Petersen* The two-fold fully conditional specification algorithm *Welch*	A model to reduce the barriers to execution of arbitrary code on potentially identifiable data without compromising privacy *Hubbard* Community Health e-Lab: Mobilising an Ethical Framework for Community-serving Uses of Individual Health Records *Ainsworth* Secure Patient Data Integration for In Silico Oncology *Coveney*	incidence and its hazard in the hypertensive population: a risk prediction function from and for clinical practice. *Alves-i-Cabratosa* National Sexual Health (NaSH) IT System in Scotland: The potential for sexual health research *McDaid* Mortality in Scottish prisoners: a cohort study *Fischbacher*	

Session 5: 1.00 -2.30 pm, Thursday 29th August						
Session 5a Obesity and BMI *(School V)* **Chair: Mark Pitman** Determination of school-based contextual factors and their association with the prevalence	**Session 5b Alcohol, smoking and drug misuse 2 *(Room 31)*** **Chair: Sheila Bird** Recurrent admissions in adolescents with victimisation-related injury: are the same adolescents also	**Session 5c Data linkage and access platforms *(School III)*** **Chair: Richard Welpton** Delivering a UK Research Platform *Thompson*	**Session 5d National Cardiovascular Registries for Monitoring Quality of Care *(Irvine Lecture Theatre)*** **Chair: Adam Timmis** Compare hospital variability in acute	**Session 5e Special Session: REporting of studies Conducted using Observational Routinely-collected Data (RECORD initiative) *(School I)*** **Chair: Sinead Langan**	**Session 5f Socio-economic inequalities in health *(Forbes Lab)*** **Chair: Hester Ward** Linking of primary care records to census data to study the association between socio-economic status	**Session 5g The value of data linkage: international perspectives *(School II)*** **Chair: Steve Pavis** Linking Healthcare Associated Infection Data to Patient Episode Data to Measure the

of overweight and obesity in children *Williams* Body mass index and coronary heart disease in the Million Women Study: a prospective study *Canoy* SurgiCal Obesity Treatment Study: using record linkage for health technology appraisal *Stewart* Body mass index and risks of haemorrhagic and ischaemic stroke in women: UK cohort linked to routine hospital discharge data *Kroll*	admitted for injury related to drug or alcohol misuse or self-harm? *Herbert* Modelling risk of smoking related disease linked to deprivation: comparison of two linked data sets. *Olajide* Record linkage validation and behavioural-risk study: drugs-related deaths soon after hospital-discharge for drug treatment clients in Scotland, 1996-2010 *Bird*	Secure Unified Research Environment: watch it work! *Churches* Development of National Linkage Infrastructure in Australia *Boyd* An approach to multi-jurisdictional wide-scale data linkage across Australia *Boyle*	myocardial infarction care and outcome between Sweden and the UK *Chung* Reporting of congenital cardiac interventional procedures: anonymous record linkage between a mandatory register and an administrative data set for capture-recapture analysis *Sims* National cardiovascular registries for measuring hospital variation in mortality after myocardial infarction: Cardiovascular disease research using Linked Bespoke studies and Electronic health Records (CALIBER) *Timmis* National Registries for Evaluating Acute Coronary Syndrome Outcomes: MINAP experience from the National Institute for Cardiovascular Outcomes Research (NICOR) *Gale*	Introduction to the RECORD initiative *Langan*/*Peterson* The value of reporting guidelines: an editor's perspective *Veitch* Extending the Cochrane risk of bias tool to cover non-randomized studies of the effects of interventions *Reeves* NHS commissioners and the use of research on observational data sets *Bardsley* Conclusion and final summary discussion *Smeeth*	and health outcomes: a nation-wide ecological study. *Alves-i-Cabratosa* Socio-economic position and childhood multimorbidity: a study using linkage between the Avon Longitudinal Study of Parents and Children and the General Practice Research Database *Cornish* Effect of socioeconomic and health inequalities on mortality: The Turin Longitudinal Study *Rasulo*	Cost of HAI to NHS Scotland *Cairns* The Kuwait Health Network: utilising routinely collected data to support quality improvement of diabetes care in Kuwait *Conway* Primary care research using national, regional and organization database record linkage in New Zealand: A foundation for international research? *Dovey* How do I love data? Let me count the ways....Using existing databases for work and health research *Koehoorn*

Session 6: 3.00-4:30 pm, Thursday 29th August						
Session 6a Interventions and health policy *(Forbes Lab)*	**Session 6b Kidney illnesses *(Room 31)***	**Session 6c Linking within families *(Irvine Lecture Theatre)***	**Session 6d Examples of E-Health research 2 *(School I)***	**Session 6e Using linked data for validation 1 *(School V)***	**Session 6f Linkage methods *(School II)***	**Session 6g Data management and meta-data: systems and platforms *(School III)***
Chair: Nayha Sethi	**Chair: Angharad Marks**	**Chair: Ruth Gilbert**	**Chair: Harry Hemingway**	**Chair: Sarah Lowe**	**Chair: John Bass**	**Chair: David Oppenheim**
Transforming a natural experiment into a health intervention evaluation using linked routine data *Rodgers* Utilising health informatics to detect an unintended consequence of antibiotics policy change: Increased rates of Acute Kidney Injury (AKI) post-operatively *Vadiveloo* Unintended consequences of change to antibiotic policy for patients with sepsis *Patton*	Tools to support clinical care in Chronic Kidney Disease: Exploiting existing data through data linkage *Marks* Using Record Linkage of Routine Health Data to address Lithium Renal Safety: Analysis of longitudinal data in a Random Coefficient Model with estimated Glomerular Filtration Rate (eGFR) *CLos* Developing a virtual population based cohort to study Chronic Kidney Disease: GLOMMS-II (the second Grampian Laboratory Outcomes Morbidity and Mortality Study) *Marks*	Comparing Relational and Graph Databases for Pedigree Datasets *Kirby* Using linked data to identify potential bias due to missing paternal details in birth registrations *Sims* Generation of Family Units in the SAIL Databank *Tingay*	Blood pressure and the initial presentation of twelve cardiovascular diseases: a CALIBER study *Hemingway* Prevalence and treatment of Active Asthma in Scotland using the Prescribing Information System *Steiner* Prevalence, management and healthcare burden of irritable bowel syndrome in Scotland *McTaggart*	Exploiting record-linkage to alcohol-related hospitalisation and mortality data to quantify non-response bias in the Scottish Health Survey *Gorman* What can linkage to electronic patient records tell us about differences in help-seeking behaviour and health-related outcomes in relation to participation in observational studies? *Cornish* Accuracy of electronic health record data for ascertainment and sub-classification of stroke outcomes in large-scale epidemiological studies: a systematic review from the UK Biobank stroke outcomes group *Woodfield*	A case for matching without names: an assessment of a cross-sector health-education data linkage using limited identifiers *Clark* How important are data-specific match-weights to probabilistic linkage? *Henderson* A generalisable method to enhance observational research through linkage to electronic primary care records. *Boyd* Privacy Preserving Probabilistic Record Linkage (P3RL) - results from a pilot study using cancer registry data *Spoerri*	Handling Large Volumes of Routinely collected Data *DSILVA* The Data Appliance? It is time to get data coming to us *Thompson* Data quality and coverage assessments for the Secure Anonymous Information Linkage databank *Demmler* The care.data programme: developing a modern data linkage service for health and social care in England *Oppenheim*

				Vascular disease in women: Comparison of diagnoses in hospital episode statistics and general practice records in England *Wright*		

Session 7: 9.00-10:30 am, Friday 30th August						
Session 7a Data linkage approaches and evaluating linkage quality *(School III)* **Chair: Katie Harron** Assessing and documenting bias and error in the linkage of Avon Longitudinal Study of Parents and Children participants to their secondary care records. *Boyd* Evaluating bias due to linkage error in anonymised linked data *Harron* Matching using	**Session 7b Integrating and curating heterogeneous clinical data, and transforming them into research-ready variables *(Forbes Lab)*** **Chair: Spiros Denaxas** Recording of acute myocardial infarction events in primary care, hospital admission, disease registry, and national mortality records *Herrett* The CALIBER Platform: phenotyping raw linked Electronic Health Record (EHR)	**Session 7c: Special Session Randomised trials utilising electronic health records *(School II)*** **Chair: Liam Smeeth** Electronic health records may offer an ideal platform to help undertake randomised intervention trials. This special session will focus on on-going and planned work involving the four HeRCs including issues in practical implementation and examples of such trials.	**Session 7d Health-care costs and provision *(School V)*** **Chair: Ronan Lyons** Outcomes following Primary knee replacement *Walters* Social inequality and hospitalisation costs for the first year of life of preterm infants *Cannon* Not all high users are created equal: Correlates of higher-than-expected health care services use in British Columbia, Canada *McGrail*	**Session 7e Enhancing routine health data *(School I)*** **Chair: Chris Dibben** Optimizing the identification of related episodes in routinely collected, non-personalized inpatient data based on record linkage *Endel* Development of an algorithm to identify and describe retroperitoneal lymph node dissections in hospital episode statistics *Evison* Methods for identifying procedural complications from	**Session 7f Challenges and solutions to the complexity of data linkage *(Irvine Lecture Theatre)*** **Chair: Mark Elliot** Overcoming methodological challenges in estimating inpatient exposure to nurse staffing by linking nursing payroll data to hospitalisation records *Schreuders* Developmental Pathways Project: Challenges for Western Australia Data Linkage *Quintero*	**Free slot**

anonymised data *Jones* Record Linkage Approach in the Dutch Biolink Project *Ariel*	data at scale for translational research *Denaxas* Automated phenotyping using free text in primary care electronic health records in patients with coronary artery disease: a CALIBER study *Shah*			national routine healthcare databases: a literature review *Keltie* Combining diagnosis, procedure, and drug information from primary and secondary care to define clinical phenotypes: a case study of Atrial Fibrillation in the CALIBER programme *Wallace*	Challenges of linking child survey data to other routinely collected data *Turner* Electronic Checking Process of Governmental Data: A Qualitative Approach *Pinto*	

Session 8: 1.00-2:30 pm, Friday 30th August						
Session 8a Special session: The role of computer science in E-health research *(Forbes Lab)* **Chair: Colin Simpson** With the ever increasing size and complexity of databases, new ways of handling and analysing 'big data' beyond traditional SQL approaches are	**Session 8b Child Health *(School I)*** **Chair: Peter Helms** The prevalence of chronic conditions in children who die: estimates based on death certificates linked to longitudinal hospital admission data *Hardelid* Childhood and Early Adulthood Predictors of Mortality: The 6-	**Session 8c Governance: Ethics and training strategies *(School II)*** **Chair: Janet Murray** Important issues with data linkage: A consensus seeking exercise *Hopf* Development of an ethics and data linkage training workshop *Flack*	**Session 8d Linkages to enhance existing data 2 *(School III)*** **Chair: David McAllister** Using a data system used to store blood results Scottish Care Information Gateway (SCI) Store in healthcare research *McAllister* Combining health, physical function and	**Session 8e Cancer research using data linkage 2 *(School V)*** **Chair: Iain Atherton** Area and individual socioeconomic factors and cancer risk: a population cohort study in Scotland *Sharpe* Socio-economic patterning in early mortality of patients aged 0-49 years	**Free slot**	**Closed meeting slot**

required. This special discussion session will provide an insight into the current problems facing clinical informaticians and discuss possible solutions.	Day Sample of the Scottish Mental Survey 1947 *Deary* SUDEP and all cause mortality in childhood onset epilepsy *Lacey* Child Medical Records for Safer Medicines (CHIMES). *Helms*	Data linkage in a federated system - opportunities and challenges *Dickinson*	social care data - a multidatabase linkage project *McGilchrist* Linkage of UK National Liver Transplant Registry and Hospital Episode Statistics: Methods and Initial Validation *Tovikkai*	diagnosed with primary bone cancer in Great Britain, 1985-2009 *Blakey* Can surveys provide a means of providing representative information on cancer survivors – a data linkage study *Atherton*		

Abstracts and special session details

This booklet contains the full abstracts of the papers that are to be presented at the conference. They are organised by session in the order that they appear in the programme. The presenting author of each paper can be identified from the main programme next to the title of the paper.

Tha majority of sessions consist of traditional paper presentations with a mixture of three or four papers in each. In addition, there are a number of special sessions which are designed to be less formal and more discussion based than traditional paper sessions. In the main programme, these sessions are labelled as:

- Session 5e “Special Session: REporting of studies Conducted using Observational Routinely-collected Data (RECORD Initiative)”
- Session 7c “Randomised trials utilising electronic health records”
- Session 8a “The role of computer science in E-health research”

More details about the programme, timetable and content of each of these special sessions can be found in the next few pages followed by the full paper abstracts.

The programme also contains a number of poster presentations. Abstracts for these can be found at the back of the booklet. The posters will be displayed for the duration of the day on Thursday 29th August in Upper College Hall with presentations scheduled for the lunchbreak.

Special Session 5e: REporting of studies Conducted using Observational Routinely-collected Data [RECORD] initiative

No reporting guidelines exist for studies based on routinely-collected health data. This session will explore specific issues relating to reporting research based on routine data, not currently covered by STROBE

Chair

Dr Sinéad Langan, NIHR Clinician Scientist, London School of Hygiene and Tropical Medicine

Programme

1.00-1.05pm: *Introduction*
Dr Sinéad Langan

1.05-1.25pm: *Introduction to the RECORD initiative*
Dr Irene Petersen, Senior Lecturer in Epidemiology and Medical Statistics, University College London and Dr Sinéad Langan

1.25-1.40pm: *The value of reporting guidelines: an editor's perspective*
Dr Emma Veitch, Senior Editor, PLOS ONE

1.40-1.55pm: *Extending the Cochrane risk of bias tool to cover non-randomized studies of the effects of interventions*
Professor Barnaby Reeves, Professorial Research Fellow in Health Services Research, Bristol Heart Institute and co-Director of the Clinical Trials and Evaluation Unit, Bristol NIHR Biomedical Research Unit for Cardiovascular Disease

1.55-2.10pm, *NHS commissioners and the use of research on observational data sets*
Dr Martin Bardsley, Director of Research, Nuffield Trust

2.10-2.30pm: *Final summary discussion*
Dr Liam Smeeth, Professor of Epidemiology, London School of Hygiene and Tropical Medicine

Special Session 7c: Utilising electronic health record systems to facilitate randomised intervention trials

Chair: Liam Smeeth

Electronic health records may offer an ideal platform to help undertake randomised intervention trials. This session will focus on on-going and planned work involving the four HeRCs including issues in practical implementation and examples of such trials.

Iain Buchan (HeRC North/N8)
The Health eResearch Centre (Farr Institute in North England) is working on methods and technologies to improve the design and conduct of clinical trials in the following areas: feasibility assessment using integrated NHS data sources; semi-automated recruitment while preserving consent for consent; and enabling patient reported outcomes using mobile technologies.

Helen Snooks (HeRC Wales/CIPHER)
This talk will cover the work of Cipher and will focus on two cluster randomised trials in which linked anonymised routine data outcomes were included. Specific issues relate to:
- consent
- permissions
- data management with both identifiable and anonymised data items
- quality of data
-matching rates
-data completeness

Tom McDonald/Isla Mackenzie
“How to get better data on medicines".
This talk will describe various e-Health trial related activities ongoing or planned in Scotland.

Tjeerd van Staa/Emily Herret (CPRD and HeRC London/CHAPTER)
Individual and cluster randomised studies within electronic health record databases.

Special Session 8a: Computing futures: Storage and management of linked data, NoSQL and I2B2 - challenges and solutions

Chair: Colin Simpson

With the ever increasing size and complexity of databases, new ways of handling and analysing 'big data' beyond traditional SQL approaches are required. This session will provide an insight into the current problems facing clinical informaticians and discuss possible solutions. The format of the session will be a panel discussion and with the aim to provoke discussion and to provide opportunities for audience members to ask questions of the panel which will consist of a number of experts from the discipline of computer science.

Linking data to build an evidence base- The Developmental Pathways Project

Glauert, R, Telethon Institute for Child Health Research
Stanley, F, Telethon Institute for Child Health Research

The Developmental Pathways in WA Children Project is a landmark project taking a multidisciplinary approach to investigate the pathways to health and wellbeing, education and juvenile delinquency outcomes among Western Australian children and youth. To achieve this, researchers from the Telethon Institute for Child Health Research and the University of Western Australia have been working in collaboration with 13 state government departments, including the WA Departments of Health, Education, Child Protection, Corrective Services, Communities, Indigenous Affairs, Treasury and Finance, Housing, Attorney General, Training and Workforce Development, Disability Services Commission, Mental Health Commission, and WA Police. The project has established the process of linking together de-identified, longitudinal, population-based data collected and stored by a large number of the WA government departments and the Telethon Institute, to create a world class research and policy planning/evaluation resource. The data are being used to evaluate the impact of policy changes, program evaluation, and risk and protective factors amongst families, individuals and communities. This presentation will provide an outline of the Project; demonstrate best practice for large, cross government linkage projects; discuss the importance of using administrative data in policy making; and provide an update on some of our research findings.

Corresponding author email: rglauert@ichr.uwa.edu.au

Comparison of incidence trends in victimisation-related injury admission in children and adolescents, between England and Scotland

Gonzalez-Izquierdo, A, Centre for Paediatric Epidemiology and Biostatistics, UCL Institute of Child Health
Cortina-Borja, M, Centre for Paediatric Epidemiology and Biostatistics, UCL Institute of Child Health
Woodman, J, Centre for Paediatric Epidemiology and Biostatistics, UCL Institute of Child Health
Mok, J, Lothian University Hospitals Division
McGhee, J, School of Social and Political Science, University of Edinburgh
Taylor, J, NSPCC Centre for Learning in Child Protection, The University of Dundee
Parkin, C, Centre for Paediatric Epidemiology and Biostatistics, UCL Institute of Child Health
Woolley, A, Centre for Paediatric Epidemiology and Biostatistics, UCL Institute of Child Health
Gilbert, R, Centre for Paediatric Epidemiology and Biostatistics, UCL Institute of Child Health

Background

Current policies to support vulnerable children in the community differ by country. Given that England and Scotland have similar healthcare systems, we compared changes over time in admission rates for victimisation-related (VR) injury in children admitted to NHS hospitals in England and Scotland.

Methods

We analysed hospital administrative data of injured children less than 19 years of age admitted to NHS hospitals in England (HES) and Scotland (SMR). VR admissions were identified by a pre-defined cluster of ICD10 codes (reflecting maltreatment, assault, undetermined cause and adverse social circumstances) in any record of longitudinally linked hospitalisation histories. We conducted a time series analysis on monthly incidence rates between 1997 and 2011, using Poisson and negative binomial regression models with a random change point.

Results

VR injury was similarly distributed across age in both countries, with the vast majority presenting in adolescents (age 11-18 years), with 78.7% (81,977/104,176) in England and 87.9% (14,860/16,909) in Scotland. The annual incidence of VR injury in infants was stable for both countries but increased in recent years by approximately 6% in England, and over 17% in Scotland, e.g.: 101.1 (95%CI 92.1, 110.9) per 100,000 child year (cy) in England and 85.6 (95%CI 63.2, 116.0) per 100,000 cy in Scotland for 2011. Trends of VR injury in children aged 1-10 years were stable and similar in both countries up to the mid-2000's, but decreased annually by 10% in Scotland from early 2006 and increased by 10% in England from late 2007. The main between-country difference was for adolescents with rates in Scotland twice as high as in England but decreasing steeply from late-2006, reaching a similar rate as in England at the end of study period - 104.7 (95%CI 100.4, 109.2) per 100,000 cy in England and 118.8 (110.0, 128.2) per 100,000 cy in Scotland. We found no evidence of diagnostic transfer to other adversity markers (self-harm or substance misuse) which were also decreasing in Scotland but less steeply.

Conclusion

We found diverging patterns of VR injury admissions in adolescents in adjacent countries with similar healthcare systems. This suggests potential effects of policy or organisational factors. Increasing trends in England are consistent with background trends for all unplanned admissions in children, but the contrast with Scotland shows that such trends are not inevitable.

Corresponding author email: arturo.gonzalez-izquierdo@ucl.ac.uk

Entering out-of-home care during childhood: Cumulative incidence study in two developed countries

O'Donnell, M, Telethon Institute for Child Health Research, University of Western Australia
Maclean, M, Telethon Institute for Child Health Research, University of Western Australia
Brownell, M, Manitoba Centre for Health Policy, University of Manitoba
Sims, S, Telethon Institute for Child Health Research
Gilbert, R, MRC Centre of Epidemiology for Child Health, UCL

In Australia and Canada, like many other developed countries, increasing numbers of maltreated children are being placed in out-of-home care and placements are occurring at an earlier age. Both factors increase the overall proportion of children experiencing out- of-home (OoH) care, along with the number of years children spend in care and the associated costs. Monitoring of these changes requires analyses that reflect changes in the timing of entry into care for cohorts of children over time instead of the usual cross sectional estimates of the annual incidence of placement in OoH care reported by government agencies.

We used linked longitudinal population level data from the Western Australian (WA) Departments of Health and Child Protection and from Manitoba's Population Health Research Data to determine the cumulative incidence of OoH Care for cohorts of children over time. Survival analysis was utilised to determine the factors associated with earlier entry into OoH care.

Manitoba had a larger proportion of children entering OoH care compared to WA (8.5% vs 1.4% of the child population by age 11 years). Over time children entered OoH care at a younger age in both WA (HR=1.4, CI=1.2-1.5) and Manitoba (HR=1.6, CI=1.5-1.7). Similar factors were associated with earlier age at entry into OoH care including, socioeconomic disadvantage, Aboriginality, young maternal age, maternal hospitalisations for mental health issues, substance misuse and assault.

Cumulative incidence estimates highlight the impact of policy on the proportion of children in some sectors of the population placed in OoH care. This information is essential for considering the potential costs and benefits of alternative in-home strategies for some communities and resources required for OoH care. In addition, cumulative incidence estimates allow comparisons between countries of the impact of regional policies about timing of placement in OoH care.

Corresponding author email: melissao@ichr.uwa.edu.au

Do NEET experiences have adverse impacts on health? Evidence from Scotland

Feng, Z, University of St Andrews
Ralston, K, University of St Andrews
Dibben, C, University of St Andrews
Raab, G, University of St Andrews

Young people who are not in employment, education or training (NEET) are a disadvantaged group and their experiences may have an adverse effect on later life. Therefore the NEET phenomenon has drawn considerable attention from academic researchers and policy makers in Britain and other countries. However, there have been theoretical arguments on the social and health consequences of the NEET experiences. So far, few studies have investigated the effect of NEET experiences on health and the limited empirical research has yielded mixed results. This paper aims to investigate whether experiences of being NEET have adverse effects on health in the Scottish context, where the prevalence of NEETs is persistently high in comparison with other parts of Britain. We used the Scottish Longitudinal Study (SLS), a 5.3% representative sample of the Scottish population linking together census records, vital registrations and patient records. We followed young people who were aged 16-19 in 1991 for 19 years up to 2010. We used descriptive and modelling approaches in our analysis. We focus on three health outcomes: limiting long term illness, hospitalisation, depression and anxiety, which are separately derived from the census, Scottish Morbidity Records, and prescribing information system. In the study we control for a number of individual and household variables from the 1991 and 2001 censuses. This research contributes to the literature on effects of lifecourse events on later health outcomes and has considerable policy implications.

Corresponding author email: zf2@st-andrews.ac.uk

Establishing a National Neonatal Research Database from operational NHS electronic records

Statnikov, Y, Neonatal Data Analysis Unit (NDAU), Imperial College London
Modi, N, for the Neonatal Data Analysis Unit Steering Board and the Medicines for Neonates Investigator Group*

Background: Over the last three decades neonatal critical care services in England and Wales have moved progressively towards the capture of a uniform clinical dataset as part of an electronic patient management system. Data include demographic details, daily records of interventions and treatments throughout the neonatal in-patient stay, information on diagnoses and outcomes, and follow-up health status at age two-years.

Aim: We aimed to establish a National Neonatal Research Database (NNRD) holding operational clinical information captured in the course of care and to make this available for multiple purposes including health services evaluations and research.

Methods: The Neonatal Data Analysis Unit (NDAU) was established in 2007 as an academic unit of Imperial College London. The NDAU sought permission from the Caldicott Guardians and Lead Neonatal Clinicians of each NHS Trust and Welsh Health Board providing a neonatal critical care service to receive clinical data from the authorised company hosting the electronic records. Regulatory approvals were obtained from the National Research Ethics Service and the National Information Governance Board to create the NNRD.

Results: Operational clinical data are received from 181 of 183 neonatal units in England and Wales. Data from over 4.8 million days of care for 399,939 neonatal admissions (347,588 babies), representing approximately 8% of all live births in England & Wales between 2006 and 2012 have been received to date and entered into the NNRD. At the NDAU, data management, using SAS 9.2 and Microsoft SQL Server 2008, includes identification of duplicate entries, evaluation of completeness of key variables, and consistency checks. The NNRD has been used for health services and clinical research, regional and national quality improvement programmes, service evaluations and national audit. Users include professional organisations, Department of Health, neonatal networks, NHS commissioners, and academic research groups.

Discussion: Neonatal services are required to provide data for multiple purposes. The availability of the NNRD has eliminated the need for repetitive collection and processing of clinical information, provided consistency in data management, and is a national resource to meet a multitude of requirements. As a repository of comprehensive clinical information covering a geographically defined population, the NNRD is a unique national and international resource.

**Members of the NDAU Steering Board and Medicines for Neonates Investigator Group: Abbott J, Ashby D, Brocklehurst P, Costeloe K, Draper E, Goldacre M, Kemp J, Majeed A, Modi N, Petrou S, Wilkinson A, Young A*

This abstract presents independent research funded by the National Institute for Health Research (NIHR) under its Programme Grants for Applied Research Programme (Grant Reference Number RP-PG-0707-10010). The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.

Corresponding author email: y.statnikov@imperial.ac.uk

The UK Neonatal Collaborative - Necrotising Enterocolitis Study: a prospect.ive population-based study using the National Neonatal Research Database

Battersby, CWS, Imperial College London
Santhakumaran, S, Imperial College London
Modi, N, Imperial College London

Background: The design of high quality clinical trials requires reliable baseline measures. The National Neonatal Research Database (NNRD) holds operational clinical information on infants admitted to neonatal units in England and Wales. Contributing neonatal units are known as the UK Neonatal Collaborative. These data can benefit research in relatively rare but serious conditions such as necrotising enterocolitis (NEC), an acute inflammatory bowel condition that predominantly affects extremely preterm neonates “32 weeks gestation and has a high mortality and morbidity rate. Most incidence data come from small studies using varying case-definitions; the UK incidence is unknown. Preventive strategies remain elusive and although enteral feeding practices are widely believed to influence susceptibility, these have not been adequately tested in randomised controlled trials (RCT). A robust evidence-based case-definition for NEC, baseline incidence data, and knowledge of current practices, are prerequisites for the design of future RCTs to test hypotheses related to feeding.

Aims: To utilise the NNRD to 1) establish an objective case-definition for NEC for national and international use; 2) determine the incidence of NEC in babies “32 weeks in a large geographically defined population; and 3) identify enteral-feed related antecedents of NEC, to inform the design of a future interventional RCT.

Methods: Data comprising 38 variables will be extracted from the NNRD for babies admitted to neonatal care over a 28 month period between 2012 and 2014. An objective case-definition for NEC will be developed from clinical and radiological signs that best predict a gold-standard definition, confirmation of NEC at laparotomy or post-mortem. The population incidence of NEC will be derived using this case-definition. The outcome (NEC or no NEC) will be compared between groups of infants with different feed exposures to investigate feed-related antecedents.

Results: Of 171 neonatal units in England,149 (87%), known as the UK Neonatal Collaborative-NEC Group, are participating in this study. In 2012, 73,622 babies were admitted to these units, of which 10,214 were born “32 weeks gestation; 309 babies had NEC confirmed at surgery or post-mortem, providing an estimated population incidence of the most severe form of NEC of 3%. Median completeness for data extracted from the NNRD for analyses ranged from 85-100%.

Discussion: In this on-going study we demonstrate that it is feasible to utilise electronic patient information captured as part of routine neonatal care from a large population to derive baseline measures to support the design of clinical trials, but that measures to maximise data quality and completeness are essential.

**Members of the NDAU Steering Board and Medicines for Neonates Investigator Group: Abbott J, Ashby D, Brocklehurst P, Costeloe K, Draper E, Goldacre M, Kemp J, Majeed A, Modi N, Petrou S, Wilkinson A, Young A*

This abstract presents independent research funded by the National Institute for Health Research (NIHR) under its Programme Grants for Applied Research Programme (Grant Reference Number RP-PG-0707-10010). The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.

Corresponding author email: c.battersby@imperial.ac.uk

Risk factors for infant deaths among singleton babies born at term in England, 2005-07

Dattani, N, City University London
Macfarlane, A, City University London

Introduction

Over 90 per cent of the live singleton births that occur in England are born at term 37-41 weeks. Although the infant mortality rate for this group of babies is only 1.8 per 1000 live births compared with the overall infant mortality rate 4.9 per 1,000 live births, they account for 40 per cent of infant deaths among live singleton births.

Ethnic origin is not recorded at registration. Therefore all analyses to date based on national data have used the mother's country of birth. But with the introduction of a central system for allocating the National Health Service Number for Babies (NN4B) in 2002, it was possible to access information on ethnicity and gestational age for all births. Information on parity is collected on a separate hospital admission system. Therefore selected data items from both of these systems were linked to birth and infant death registration records in England to enable analysis of infant deaths by gestation, birthweight, ethnicity, parity, mother's age, marital status and area deprivation score.

Method

All records of singleton live births linked to NN4B and hospital records in England for 2005 to 2007, for babies born at term and for babies whose ethnic group ethnic group was recorded as White British, Pakistani, Bangladeshi, Indian, Black African and Black Caribbean were used.

Odds ratios and p-values for the univariate and multivariate analysis were derived from logistic regression in SPSS.

Results

There were 2,798 infant deaths among 1,510,376 singleton live births in England from 2005 to 2007, giving a rate of 1.85 per 1,000 live births. Infant mortality for babies of Pakistani ethnicity was significantly higher at 3.14 per 1,000 live births compared to white British, Indian, Bangladeshi, black Caribbean and black African babies. This rate remained significantly higher at 2.54 per 1,000 after controlling for mothers age, birthweight, parity, marital status and area deprivation score.

Conclusion

Infant mortality for babies of Pakistani ethnicity remains high at term compared to all other ethnic groups in England, after controlling for socio-demographic factors and area deprivation score. These babies have high mortality rates due to congenital anomalies and these may be attributed to autosomal recessive inheritance.

Corresponding author email: n.dattani.1@city.ac.uk

Creating the Scottish Congenital Anomalies Linked Dataset - A New Methodology

Nolan, J, Information Services Division, NHS Scotland, Edinburgh, UNITED KINGDOM
Morris, C, Information Services Division, NHS Scotland, Edinburgh, UNITED KINGDOM
Clark, D, Information Services Division, NHS Scotland, Edinburgh, UNITED KINGDOM

Aim:
To create a Scottish congenital anomalies linked dataset using "linkable" chi-seeded maternity and neonatal datasets. This is a new approach for producing the data which no longer requires creating a single probability matched linked maternity and neonatal data file. In the absence of a comprehensive congenital anomalies register for Scotland this dataset is required to record all congenital anomalies detected at birth or during infancy. Results from the dataset have been published in the Scottish Perinatal and Infant Mortality and Morbidity Report for 2011.

Background:
Due to the expansion of the use of CHI (Community Health Index) number as an identifier across the National Health Service in Scotland, a more innovative approach to linking records from different datasets is now possible. All datasets within ISD, containing sufficient personal identifiers, are now seeded with the CHI number improving their "linkability". This year the congenital anomalies linked dataset for Scotland was created for the first time using these separate chi seeded maternity and neonatal datasets.

Method:
The congenital anomalies linked dataset is produced by linking routine national data including the Scottish Birth Record, SMR11 neonatal discharge records, Stillbirth and Infant Death records, SMR02 delivery records and SMR01 hospital admission records. Each of these stand alone datasets is now seeded with either the mother or baby chi number. In addition, NRS (National Records for Scotland) birth registration records have been seeded with both the mother and baby chi and are used as a "spine" for linking the different record types together.

Results:
The results from the congenital anomalies linked dataset are very close to those from the previous methodology. For example, total reported numbers of Down's Syndrome for the years 2006 to 2009 were previously 248, increasing by 2% to 254 under the new methodology. The equivalent change for Spina Bifida was an increase of 1%.

Conclusions:
Work to better understand the differences between outputs from the new "linkable" datasets and the old probability matched linked file is ongoing, however the results from the congenital anomalies dataset confirm the validity of the new chi based approach for linking maternity and neonatal records. In addition, this methodology ensures that future linkages of maternity and neonatal data to other ISD held data, and to external data, will be simpler, and adheres to SHIP principles by limiting the risk of disclosure.

Corresponding author email: john.nolan@nhs.net

Use of multiple electronic data capturing systems to monitor patients at a major trauma centre

Bhatti, J, Health Informatics Department, University Hospitals Birmingham NHS Foundation Trust, United Kingdom
Nisar, S, Health Informatics Department, University Hospitals Birmingham NHS Foundation Trust, United Kingdom
Evison, F, Health Informatics Department, University Hospitals Birmingham NHS Foundation Trust, United Kingdom
Begaj, I, Health Informatics Department, University Hospitals Birmingham NHS Foundation Trust, United Kingdom

University Hospitals Birmingham NHS Foundation Trust (UHB) was awarded Major Trauma Centre status in March 2012, making it the first port of call for people with life-threatening injuries in the region as well as injured soldiers stationed abroad.

Patients with major trauma are those with serious, multiple injuries that require 24 hours a day and 7 days a week emergency access to a wide range of clinical services and specialist staff. For example, doctors may be required to attend to a patient with head and neck injuries, chest, pelvis and other bone fractures.

UHB utilises data from various systems for analysis and service redesign, for example staffing levels have been adjusted due to the number of major trauma patients arriving out of hours in A&E.

An internally developed web based system refreshes at regular intervals updating the list of major trauma patients which are flagged in the A&E system. This data is linked with the inpatient system to extract the patient's current location and to establish other details such as using the patient or GP's postcode to establish the nearest district hospital. This data provides the basis of the patient's details allowing various clinicians and administrators to add rehab, injury and communication details.

The interfaces with the various hospital systems helps to ensure all the patients are captured for commissioning purposes, and more importantly to increase quality of care, in line with the Government's quality based agenda for the NHS. Furthermore the applications also provide data for submissions to The Trauma Audit & Research Network (TARN), a collaboration of hospitals across the UK and parts of Europe to assist with research.

Daily reporting tools have been created from the database allowing analysis of activity trends, including: arrival modes (self present/ ambulance/helicopter), injury mechanisms and length of stay, enabling better management of a patient's stay.

Further enhancements of the tool are planned, including automatic data extractions for clinical information such as injury severity scores and notifications from a Neurosurgery system, when j trauma patients are transferred from tertiary providers. It is envisioned that this system could be used nationally for the basis of data collection for all major trauma events.

Corresponding author email: Felicity.Evison@uhb.nhs.uk

Antidepressant prescriptions and subsequent health service utilization in Austria: A record linkage study in a country with a fragmented payment system and only partially available unique patient identifiers

Katschnig, H, Ludwig Boltzmann Institute for Social Psychiatry, Vienna; IMEHPS.reserach, Vienna, Austria
Endel, F, Technical University of Vienna, Austria
Endel, G, Main Association of Austrian Social Security Institutions, Vienna, Austria

Austria, a federal country with 8.4 million inhabitants and 9 provinces, has a fragmented health care provider payment system with 19 different mandatory social health insurance institutions covering 98% of the population. Service utilization is recorded in many different databases, with different structures and semantics, and only partly reported with a unique patient identifier. In an ongoing project a countrywide database (GAP-DRG) is being established (with the years 2006 and 2007 already finished) in which service utilization records are linked after pseudonymisation of the UPI and by special matching methods. The database covers the whole range of health services, including hospital, specialized outpatient and primary care. In an ongoing study the cohort of patients for whom an antidepressant prescription (ATC N06A) was filled in 2006 was followed up for 12 months in order to identify the patients' subsequent health service utilization in the primary care, specialized outpatient and in-patient sector. In 2006 4,146.848 prescriptions for antidepressants were filled, thereof 1,083.218 by 424.281 patients in the last quarter of 2006. 19.264 persons who died before the end of 2007 were excluded from follow-up, leaving a cohort of 404.657 patients to be followed up. In nearly 4 of 5 patients the prescription came from a general practitioner (who has no gatekeeping function in Austria). Preliminary follow-up results show that only around 3% of the cohort was admitted to a psychiatric hospital bed, but ten times as many to a non-psychiatric hospital bed. 18% saw a psychiatrist in an outpatient service, but over two thirds attended a non-psychiatric specialized out-patient service. Co-prescriptions of medication for physical disorder are currently analysed in order to determine whether physical comorbidity (elicited by co-prescription) is responsible for the high rate of non-psychiatric service utilization. Full results will be reported at the conference.

Corresponding author email: heinz.katschnig@meduniwien.ac.at

The effect of primary care psychologist treatment on health care utilization in general practice. A longitudinal data linkage study.

Prins, MA, NIVEL, Netherlands Institute of Health Services Research, Utrecht, the Netherlands
Verhaak, PFM, NIVEL, Netherlands Institute of Health Services Research, Utrecht, the Netherlands
Verheij, RA, NIVEL, Netherlands Institute of Health Services Research, Utrecht, the Netherlands

Objective. Literature suggests that serious mental health problems enhance the use of health services and that psychological interventions can reduce this effect. In this study we investigate whether this effect is also found in primary care patients with mild to moderate mental health problems.

Methods. Routine electronic health records from a representative sample of 128 general practices were linked to health records from 150 primary care psychologists (PCPs), participating in the NIVEL Primary Care Database. Statistics Netherlands acted as a trusted third party by making it possible to link the data on the basis of date of birth, gender and four-digit postal code. Data were enriched with information from the population registry and other sources available at Statistics Netherlands. We analyzed the number of contacts, health problems presented and prescribed medication in general practice six months before and six months after the PCP treatment using poisson and logistic multilevel regression analysis for repeated measures.

Results. Of the 43,899 patients who were treated by 651 PCPs in 2009 and 382,726 patients who were registered in one of the 128 GP practices in 2009 that could be linked to the population registry, 1356 patients were found in both networks. For 503 patients complete GP data were available on contacts, diagnoses and prescriptions in 2008, 2009 and 2010. Nearly all 503 patients consulted their GP during the six months preceding the PCP treatment (97.8%) and also in the six months after the PCP treatment had ended (95.2%). The frequency of consultations was higher before than after the PCP treatment (6.1 vs 4.8). Significantly fewer patients consulted GPs specifically for psychological or social problems (46.3% vs 38.8%) and fewer patients had psychotropic drugs prescriptions (26.8% vs 22.7%) after PCP treatment.

Conclusion. This study shows that the use of GP services decrease after psychological intervention in primary care, at least in the short run. Although differences are statistically significant, consultation rates are still relatively high. At this point, we cannot say whether the effect is still there in the long run or whether the effect increases or decreases in time. Since the NIVEL Primary Care Database is being expanded at this time, future studies will be based on a greater number of persons over a longer period of time.

Corresponding author email: m.prins@nivel.nl

Body mass index and hospital admissions in UK women: a prospective cohort study

Reeves, GK, Cancer Epidemiology Unit, University of Oxford
Balkwill, A, Cancer Epidemiology Unit, University of Oxford
Cairns, BJ, Cancer Epidemiology Unit, University of Oxford
Green, J, Cancer Epidemiology Unit, University of Oxford
Beral, V, Cancer Epidemiology Unit, University of Oxford

Adiposity is associated with many adverse health outcomes but little direct evidence exists about its impact on use of health care services. We examined the relationship between body mass index (BMI) and hospital admission rates in a large UK cohort of middle-aged women. Hospitalisation rates by BMI were estimated, standardised for age, region of recruitment, socioeconomic status, reproductive history, smoking status, hormonal therapy use and alcohol intake. Proportional hazards models were used to estimate adjusted relative risks of hospitalisation separately for 25 common types of admission. Over 1.2 million women aged 50-64 at entry were included in analyses. During an average follow-up of 9.2 years, there was a total of around 2.8 million incident hospital admissions. We observed significant increases in the risk of admission with increasing BMI for all hospital admissions, and for 19 of the 25 types of hospital admission considered here. The largest associations with BMI were for admissions with diabetes, knee-replacement, gallbladder disease and venous thromboembolism.

Corresponding author email: gill.reeves@ceu.ox.ac.uk

Are psychotic experiences pathological? Comparison between information on self reported and observer-rated psychotic experiences in the ALSPAC birth cohort study and clinical information in linked primary care records

Davies, AJV, University of Bristol
Zammit, S, University of Cardiff
Sullivan, S, University of Bristol
Macleod, JM, University of Bristol
Cornish, R, University of Bristol
Boyd, AW, University of Bristol

Background

Psychotic experiences (PEs) are reported by around 15% of the general population. Several epidemiological studies, for example those investigating possible consequences of adolescent cannabis use, have used such symptoms as a proxy for risk of clinical psychosis. We have previously reported that PEs in the ALSPAC cohort, though predicted by cannabis use, are associated with established psychosis risk factors. Other studies have shown PEs predict psychotic disorder, but with low positive predictive value. It has also been seen that risk factors for PEs are not specific, for example they also increase the risk of depression. This calls the validity of PEs as a proxy for clinical psychosis into question. We investigated this question of the meaning of PEs by relating self-reported psychotic experiences in the Avon Longitudinal Study of Parents and Children (ALSPAC), a large birth cohort, to information in the primary care records of ALSPAC participants.

Methods

This study looks to gain a clearer understanding of PEs using primary care measures and the ALSPAC cohort. Data collection within the cohort continues primarily through questionnaires, but also clinics and linkage to routine data. Subjects eligible to be a part of the cohort have been linked to the General Practice Research Database (GPRD) through the NHS Information Centre (NHS IC). GPRD is an anonymised database of primary care records of approx. 5 million patients within the UK (approx. 5% national coverage). Using ALSPAC data on PEs and associated variables such as friendships, relationships and substance use, we look to determine associations with primary care measures. GP-recorded PEs will initially be explored; it is thought that existence of these records within GPRD will be extremely low, if not non-existent. The second part of the analysis will look to determine patterns of characteristics, within primary care measures, amongst those with reported PEs. Some of the measures to be explored will include service use within the health care system, records of sickness certification, medication and re-occurring GP-diagnoses amongst those with self reported and observer rated PEs. Logistic regression models will be used to analyse the data.

Results

Within the linked GPRD data approximately 40% had PEs data available at age 13. When looking at self reported measures 24.14% were reported as having definite PEs within GPRD. Coverage of those with suspected or definite PEs, using observer rated measures, within the database, stands at 11.52%. Analysis for this project is ongoing and final results looking at associations will be presented.

Conclusions

The limitations of this study will be discussed as well as the final results and what they may suggest. Further analysis plans and ways of improving and overcoming limitations will also be presented.

Corresponding author email: a.davies@bristol.ac.uk

Psychotropic medication utilisation in care home residents age 65 or older compared with the equivalent general population in Scotland

Stewart, T, College of Pharmacy, University of Kentucky
McTaggart, SA, Information Services Division, NHS National Services Scotland

OBJECTIVE: To compare psychotropic medication utilisation in the elderly care home population with the equivalent non-care home population including comparison by gender and age band between and within the two populations

METHODS: Information on psychotropic medication prescribing (anxiolytics, hypnotics and oral antipsychotics) during 2011 was extracted from the national primary care prescribing information system (PIS) for all patients aged 65 years or older on 1 January 2011. Information was collected on the type of medication, number of prescriptions, number of defined daily doses (DDD) along with patient demographic information including gender, age and care home residency status and NHS Board.
Data were analysed by care home residency status and type of medication received and were stratified by age and gender. Analyses included percent of the patient population exposed and relative risk.

RESULTS: A total of 879,492 people were included, 32,372 of whom resided in care homes (3.7%). Males comprised 28% and 43% of the care home and non-care home populations, respectively and age groups 65-74, 75-84, and 85+ made up 12%, 35%, and 53% of the care home population, respectively, while making up 55%, 34%, and 11% of the non-care home population, respectively. More care home residents were exposed to psychotropics than non-care home residents (41.6% vs. 12.1%, $p<0.001$) and to each class of psychotropic: anxiolytics 17.0% vs .5.2%; hypnotics 16.1% vs. 6.8% and antipsychotics 23.2% vs. 1.8%. Males were more likely to receive psychotropics than females in care homes (43.6% vs. 40.9%, $p=0.048$), but less likely to receive psychotropics in the non-care home setting (9.4% vs. 14.3%, $p<0.001$). Treatment with psychotropics was more common for every age group living in the care home (65-74 years RR=4.6 95% CI=2.93-6.29, 75-84 years RR=3.4 95% CI=2.42-4.54, 85+ years RR=2.4 95% CI=1.31-3.47). The percent of non-care home patients treated with psychotropics increased as patients aged (65-74 years 10.7%, 75-84 years 13.4%, 85+ years 15.9%). Conversely, the percent of care home patients treated with psychotropics decreased as patients aged (65-74 years 49.4%, 75-84 years 46.0%, 85+ years 37.4%).

CONCLUSION: Elderly residents of care homes have increased exposure to psychotropic medications compared to the equivalent non-care home population. Whether such treatment is initiated in the care home setting and whether the use of these drugs is clinically appropriate in a frail, elderly population warrants further investigation

Corresponding author email: stuart.mctaggart@nhs.net

Can the use of routine data enhance collection of the primary outcome in the SHIFT Trial

Wright-Hughes, A, University of Leeds, Leeds Institute of Clinical Trials Research & the School of Medicine, Clinical Trials Research Unit
Graham, L, University of Leeds, Leeds Institute of Clinical Trials Research & the School of Medicine, Clinical Trials Research Unit
Farrin, A, University of Leeds, Leeds Institute of Clinical Trials Research & the School of Medicine, Clinical Trials Research Unit
Cottrell, D, University of Leeds, School of Medicine

Having obtained routinely available data from the NHS Health and Social Care Information Centre (HSCIC), we explored the reliability of this data compared to data collected directly from hospital records, and evaluated the benefits and limitations of its use in the SHIFT Trial.

SHIFT is a randomised controlled trial of family therapy vs. treatment as usual for adolescents following self-harm. The trial’s primary endpoint, repetition of self-harm leading to hospital attendance, requires timely and regular collection of hospital attendance data to inform the timing of analysis, and is thus a resource-intensive task requiring researcher visits to many hospitals across England to interrogate local medical records.

Hospital Episode Statistics (HES) data held by the HSCIC contains information provided periodically by English hospitals. The main use of this data is to provide England-wide statistics to inform frontline decision makers. We assessed whether this is a complete, accurate and reliable means of acquiring outcome data for the SHIFT trial.

Our logic was, should the data be reliable, benefits of this approach would be: a) Regular, fast England-wide data retrieval rather than collection from an identified, limited hospital “pool”; b) avoidance of potentially biased data collection due to more frequent visits to some hospitals; and c) freeing up researcher resources.

Data retrieved from the HSCIC was compared to data previously gathered by trial researchers from pre-identified Hospitals, with consideration given to:

a) Linkage rates
b) Number and proportion of episodes retrieved via the HSCIS compared to those previously identified by researchers;
c) Reliability of both Accident and Emergency attendances and hospital admissions;
d) Percentage of self-harm episodes recorded & coded appropriately in HES;
e) Percentage of required data items retrieved for each episode from HES;

Our comparison found advantages of data collection via the HSCIC to include the acquisition of more comprehensive and timely trial outcome data, potentially at a reduced cost, whilst disadvantages included ambiguity in the classification of self-harm relatedness for a proportion of episodes.

Overall, advantages were found to far outweigh disadvantages to the SHIFT trial, and thus a change to the method of primary outcome data collection was instigated. As such our resulting data collection strategy allows for the identification of hospital attendances via the HSCIC with further data collection via targeted researcher visits to hospitals for episodes requiring supplementary information.

Findings relating to our comparison of data collection methods and the reliability of this data will be presented.

Corresponding author email: a.wright-hughes@leeds.ac.uk

Results of depression screening in large community based population

Purves, D, Robertson Centre for Biostatistics, University of Glasgow
Jani, B, General Practice & Primary Care, University of Glasgow
Barry, S, Robertson Centre for Biostatistics, University of Glasgow
Mair, F, General Practice & Primary Care, University of Glasgow
McLean, G, General Practice & Primary Care, University of Glasgow
Cavanagh, J, Institute of Health and Wellbeing, University of Glasgow

Aims

Depression status in a community based population, comprising patients with diagnoses of diabetes, cardiovascular disease (CVD) or stroke, was examined through the use of routinely collected patient data. The aims were to describe the incidence of new depression cases in this population, as identified through screening, explore any association between these and a range of patient characteristics and examine if patient screening impacted on commencement of treatment.

Data and methods

General practices in Glasgow and surrounding areas routinely conduct annual health assessments on subjects with chronic diseases. They are incentivised to perform depression screening using the depressive subscale of Hospital Depression and Anxiety Score; HADS-D. The Keep Well/Enhanced Services Data Group gave us permission to use this data, and ethics approval was obtained from the West of Scotland Research Ethics Service.

Patients were identified as being under-treatment and ineligible for depression screening if they were receiving antidepressants or psychotherapy. Of remaining patients, we defined positive screens for depression if a HADS-D was recorded. Regression models explored associations with age, gender, socioeconomic status (SES), comorbidities and HADS.

Results

A cohort of 125,143 patients, aged between 18 and 90, were identified as having a previous diagnosis of at least one of diabetes, CVD or stroke in 2008-09. 15,659 (12.5%) were already under treatment for depression. Of the remaining patients, 35,537 (32.5%) had a HADS recorded; 28,457 (80%) had a HADS < 8, 4,155 (12%) 8-10 and 2,925 (8%) “11.

In the screened population, females, the more socioeconomically deprived and patients with multiple morbidities were more likely to have a raised HADS. Following a raised HADS, females were 59% more likely to be treated for depression than males (odds ratio (OR) 1.59, 95% CI 1.66 - 1.72, p<0.001) and younger patients (under 75 years) up to twice as likely to be treated compared to older people (OR 2.00, 95% CI 1.58 - 2.54, p<0.001). The odds of starting new treatment increased by 11% for each unit increase in HADS (OR 1.11, 95% CI 1.07 - 1.14, p<0.001). Neither a patient’s SES nor the presence of comorbidities affected the odds of starting new treatment.

Conclusions

A minority of eligible patients had a HADS recorded, despite incentivisation. Screening detected a relatively high incidence of depression in patients with chronic diseases, particularly in females, deprived patients and those with multiple morbidities. Routinely collected data such as this provides very useful population-level information, but could be vastly improved by being more complete, possibly by targeting high risk groups.

Corresponding author email: david.purves@glasgow.ac.uk

Beyond "Big Data": assessing the quality and utility of administrative data for research use

Smith, M, Manitoba Centre for Health Policy, University of Manitoba, Canada
Lix, LM, University of Manitoba, Manitoba, Canada
Azimaee, M, Institute for Clinical Evaluative Sciences (ICES), Toronto, Canada
Hong, S, Manitoba Centre for Health Policy, University of Manitoba, Canada
Towns, D, Manitoba Centre for Health Policy, University of Manitoba, Canada
Nicol, JP, Manitoba Centre for Health Policy, University of Manitoba, Canada

As part of a large multi-year Canada Foundation for Information (CFI) grant the Manitoba Centre for Health Policy (MCHP) was awarded funding to add 15 new large and complex social, clinical, or health services data sets to its data repository. In concert, MCHP undertook to revise its data management procedures and focus attention on quality assessment and documentation of administrative data.

This presentation will focus on the Data Quality Framework and its implementation in the Data Quality Toolkit, a set of SAS macros to automate the data quality evaluation and reporting process, and the new online Data Dictionary tool which allows exploring and visualizing data and documentation. Several case studies are presented highlighting the utility of these tools in exploring data and uncovering anomalies. An advisory committee, which includes experts from clinical or government programs in conjunction with academic colleagues, is used to oversee and vet the acquisition process and the function of this committee will be discussed. Finally, an overview of a new 6-step data management process in use at MCHP and its relationship to data linkage and data quality will be provided as background information. In the process we answer the question "What's beyond Big Data".

Corresponding author email: mark_smith@cpe.umanitoba.ca

Improving linked data quality, research outcomes and reducing costs using graph theory and graph databases

Farrow, JM, Farrow Norris and SA.NT DataLink

We present a new method for managing linked record/event data by storing rich comparison data as a network or graph (in the computer science sense) in a graph database. Current record-linkage systems predominantly store data in a relational database or flat-file format concentrating on final assigned “groups” or “clusters”. The calculations made to assign such groupings are either discarded or archived in a form not readily accessible to further computation. By instead storing all data, including subsequent clerical review data, in a graph database, calculations may be revisited without the overhead of repeating past work and new functionality becomes possible.

We describe the methodology used by the Next Generation Linkage Management System (NGLMS) built for SA.NT DataLink where a graph structure is used to store a “more natural” representation: records as vertices and comparisons as edges between vertices.

This graph approach allows improvements to data handling (reduces double handling of data), data quality, resource allocation, and allows new analysis and extraction techniques to be applied and explored. Clerical review may be done on-demand (just-in-time), piecemeal, deferred, repeated, improved and even excluded (temporarily or permanently) from future calculations; entire datasets may be efficiently selected or completely ignored for computational purposes. A new clerical review process based around the graph approach will be described.

By retaining rich comparison information new data analysis techniques are made practical and some will be presented. Fast cluster detection and analysis is facilitated by storing the information in a 'native' format which allows efficient graph traversal algorithms to be used to detect clusters of related records. Modern graph databases are designed to facilitate these types of queries and the approach avoids many of the problems associated with a traditional relational approach.

Non-traditional information may also be stored in the graph and exploited: genealogical information, tribal and clan kinships structures, DNA/genetic similarities. Feedback from end-users may be attached to link quality.

The graph approach enables dynamic future functionality and link manipulation where different extractions of the underlying data may be made with differing characteristics to suit different research purposes. Extractions may be performed with different 'quality' levels, e.g. high-precision vs high-recall from the same underlying data which may be used to improve research outcomes; extractions may be performed using differing extraction algorithms (cf. a flat-file/traditional approach where only a single description of the linkage clusters is kept).

Corresponding author email: james@fn.com.au

Sampling based clerical assessment

Guiver, T, Australian Institute of Health and Welfare
Boyd, J, Curtin University

Clerical review is a time intensive stage of the data linking process, requiring a high level of visual display unit and keyboard use. Some data linkage units in Australia continue to undertake significant amounts of clerical review, employing a number of staff whose sole role is clerical review. This work is tedious, cannot be undertaken for extended periods at a consistently high level of accuracy and has attendant occupational health and safety risks.

A sampling based approach can be used to dramatically reduce the amount of clerical inspection required. This approach can replace complete inspection of record pairs in the clerical review range with a sample selected using a probability based sample design.

The sample based approach has a number of applications. Firstly, sampling can be used to assess a full clerical review. Secondly, sampling can provide an accurate and reliable means of assessing and setting the most appropriate bounds for a subsequent full clerical review. Finally, for large linkages a sample based approach can be used set a single acceptance bound; while this will introduce missed and false links, the extent can be quantified resulting valuable quality measures that can used to inform analyses.

This paper describes the results of a research project undertaken jointly by the Australian Institute of Health and Welfare and the Centre for Data Linkage at Curtin University which looked at the efficacy and benefits of adopting a sample based strategy to clerical review and makes recommendations for how to reduce the burden of manual review.

Corresponding author email: tenniel.guiver@aihw.gov.au

ESRC Welsh Government Data Linking Demonstration Projects 2012-13: Methodological Challenges and Lessons Learned

Lowe, SE, Welsh Government
Heaven, MH, Swansea University
Dolman, R, Welsh Government

The UK Economic and Social Research Council and Welsh Government (WG) jointly funded a Fellow to work jointly between WG and the Health Information Research Unit (HIRU) at Swansea University (home of the SAIL - Secure Anonymised Information Linkage - System) from October 2011 to March 2013. The Fellow has worked with WG analysts and policymakers to deliver three data linking projects designed to:

a) fill evidence gaps for policy-making;
b) demonstrate the unique contribution of data linking and;
c) identify the methodological challenges of and solutions for linking data and the analysis of linked data.

The projects all have a significant health component and two link respectively to data from housing and education. The topics were as follows i) describing and comparing patient pathway variations for patients with multiple chronic health conditions, ii) identifying the health impacts of home energy efficiency interventions and iii) identifying the heath impacts of the Wales early years intervention Flying Start.

The presentation will demonstrate a range of challenges and lessons learned - with proposed solutions where possible - in the following areas: 1) practical issues around data acquisition and information security; 2) managing expectations - how and why everything takes longer than you expect!; and 3) the methodological challenges of linking multiple administrative data sets, including an examination of the common problems encountered plus lessons learned about data quality, validation and analysis methods. As a result of lessons learned, further work is being undertaken from April 2013 to improve linked data for social care and (TBC) housing for Wales. The presentation will briefly touch on issues around using lessons learned during data linking to improve upstream data collection.

Corresponding author email: sarah.lowe@wales.gsi.gov.uk

Implementing safe effective proportionate governance in NHS National Services Scotland

Murray, J, NHS National Services Scotland
Wood, R, NHS National Services Scotland
Ruddy, P, NHS National Services Scotland
Stewart, A D, NHS National Services Scotland

The SHIP Good Governance Framework describes a three-step risk-based approach to governance using benchmark core considerations and privacy risk assessment to allocate applications for research uses of data to the appropriate level of scrutiny. This proportionate governance model has been adopted by the Scottish Government in “Joined Up Data for Better Decisions” and implementation of this approach across Scotland is seen as essential to enable efficient safe use of electronic data for research. It is in the public interest that the model is robust.

NHS National Services Scotland (NSS) is piloting criteria for proportionate governance of the research use of Scottish Morbidity Records and other national health datasets in its control. This is expected to inform a Scotland wide model. For over 20 years, NSS Privacy Advisory Committee (PAC) has enabled research, advising NSS and National Records of Scotland on the correct balance between protecting personal data and making it available for uses in the public interest, and ensuring compliance with legal and other obligations. The committee comprises mainly lay members and is supported by experts in information governance.

Twelve criteria were developed to support combined assessment of the privacy risk and the benchmark core considerations: Public Interest, Safe Data, Safe People, Safe Environment and relative risks. This involved review of previous applications, and consultation with committee members and those involved in the SHIP Governance Work-stream. Standards were established to define applications as green, amber or red on each criterion and a flow chart developed to enable allocation to appropriate scrutiny: no further review, fast track review or full review by the committee. The Privacy Advisory Committee application form was modified to capture information required for assessment with input from Research Co-ordinators who now support researchers in developing research proposals.

The pilot aims to test the robustness of the criteria in differentiating applications reliably and accurately, to modify them as necessary, and to establish a system whereby many applications can be processed by an information governance expert with further review of high risk applications only. It aims to identify the impact on workload of all stakeholders and on application processing time and to inform indicative response times within NSS.

The paper will present the proportionate governance process in detail, report the lessons learned and go on to consider what infrastructure may be required to implement a consistent approach throughout Scotland.

Corresponding author email: janet.murray1@nhs.net

Control and Trust as the Foundations of Public Acceptability: Reflections from the Public Engagement workstream of SHIP

Aitken, M, University of Edinburgh
Cunningham-Burley, S, University of Edinburgh
Pagliari, C, University of Edinburgh

The public engagement workstream of the Scottish Health Informatics Programme (SHIP) has conducted a series of public engagement activities to explore public attitudes towards SHIP, data-linkage and uses of personal data more broadly. This has involved three main components:

1) A series of eight focus groups with a diverse range of public groups across Scotland;
2) A workshop with a range of stakeholders on the topic of trust in researchers and;
3) A series of three deliberative workshop events with members of the public.

A key theme to emerge from this research has been the importance of perceived control as a factor shaping public responses to data-linkage and use. It is clear that the extent to which individuals' feel in control of their data informs the extent to which they support the aims of data linkage for health research. A second important and related theme is the importance of trust: where individuals do not trust organisations or individuals making decisions about uses of their data they do not support data-linkage or data-sharing for research purposes. Control, whilst consistently identified as a crucial consideration, can be interpreted and facilitated in many different ways and can imply varying degrees of devolution of power to members of the public. Where levels of trust are high the level of control perceived as appropriate is lower, conversely where individuals do not trust the organisations/individuals using or making decisions about their data they typically assert the need for very detailed, individual-level control over data. This has important implications for the governance of data-linkage and sharing and for the operation of SHIP and similar programmes.

In this presentation we will present a critical summary of our findings from our public engagement work and reflect on what these mean for governance of data-linkage and how control and trust might be built into the operation of SHIP and the new e-Health Informatics Research Centres in the UK.

Corresponding author email: mhairi.aitken@ed.ac.uk

Using Electronic Health Records for research purposes: key findings from a survey on patient and public attitudes

Papoutsi, C, NIHR CLAHRC for Northwest London, Imperial College London
Reed, J, NIHR CLAHRC for Northwest London, Imperial College London
Marston, C, London School of Hygiene and Tropical Medicine
Majeed, A, Imperial College London
Bell, D, NIHR CLAHRC for Northwest London, Imperial College London

This paper presents findings from a large-scale, cross-sectional survey looking at patient and public attitudes on integrated Electronic Health Records (EHRs) used for healthcare, planning and policy, and health research purposes. The questionnaire was self-completed by 5331 participants (response rate 85.5%) recruited from 16 hospital outpatient and general practice clinics in West London between August and September 2011. Of the full sample, 2554 respondents (48%) provided complete data for this analysis focusing on research-related variables.

The majority of participants (82%) responded positively to the use of EHRs for research, with 69% in favour of sharing their complete medical history if personal identifiers had been removed. Only a small proportion (13.4%) would allow their full medical history to be shared together with their personal identifiers, while slightly more (18.2%) would not support any use of their records for research at all.

When asked about their preferences for access to EHRs by different types of research groups, participants showed support for NHS and academic researchers at similar levels as for research overall. Pharmaceutical companies and health charities, however, were not equally supported, as higher proportions of respondents (44% and 33% respectively) would not allow these groups any access to their identifiable or anonymised records.

High levels of security concerns accompanied these differences in preference. Almost 80% of participants stated that they would worry about the security of integrated EHRs being used simultaneously for health provision, planning and research. Interestingly, of those worried about the security of EHRs, 57% would still support their development and almost 70% would be in favour of their anonymised records being used for research.

Older participants, people from ethnic backgrounds other than White British and those with lower educational qualifications were more likely to support access to their identifiable, rather than anonymised, records for research by all user groups. Respondents who were less satisfied with the NHS were more likely to be opposed to their records being shared for research and more worried about security risks.

These findings indicate that while most individuals have positive views about EHRs being used for research, they have specific preferences about which researchers should have access to their information. When broadening the scope of aggregated research databases in the NHS, security concerns must be addressed to retain the trust of patients and the public.

Corresponding author email: c.papoutsi@imperial.ac.uk

The Freedom of Information Act (2000) in healthcare research: A Systematic Review

Fowler, AJ, Barts and the London SMD, QMUL, London
Agha, RA, Plastic Surgery department, Stoke Mandeville hospital, Ayelsbury, Bucks, HP21 8AL
Camm, CF, New College, Oxford
Littlejohns, P, Department of Primary care and public health, KCL, London

Background

The Freedom of Information Act passed into law in 2000 and came into full force in 2005. As a result, public authorities have a legal obligation to provide information through an approved publication scheme and in response to requests. This has resulted in a boost for information transparency and a wealth of information that was previously difficult to access may now be obtained through a simple request. Our objective was to assess the use and utility of the Freedom of Information Act in healthcare research between 2005 and 2013.

Methods

An online search of the EMBASE, Medline, CINAHL, psycINFO, BNI, AMED, HMIC and Health business elite databases was conducted from January 2005 to January 2013 using the terms “Freedom of information”, “Freedom of information act”, and “Freedom of information act 2000.” The search was restricted to English language papers only. Results were manually screened for those utilising the freedom of information act in healthcare research.

Results

114 papers were identified after duplicates were removed. From this, 16 papers met the inclusion criteria for manual abstract screening. The median number of requests made was 45 (range 1-172), the total number of requests was 1,582. A total of 25,658 pieces of data were retrieved by all studies. The range of subject areas included litigation, surgery, funding, laboratory provision, pharmacological safety, problem gambling and midwifery disciplinary patterns. A median of four questions were asked per study (range 2-8) and the response rate overall was 81%. The NHS litigation authority responded to 100% of requests, while Primary Care Trusts had the lowest response rate for healthcare bodies of 69%. A negative correlation between response rate and number of requests made reached statistical significance (-0.605, p=0.013).

Conclusion

The Freedom of Information Act has provided a welcome tool to increase information transparency. Researchers should make greater use of the act to access information they need that is not otherwise disclosed. We discuss issues with research utilising the act and how future research of this type could be optimised.

Corresponding author email: ha09410@qmul.ac.uk

The Candida in Pregnancy Study (CiPS): a randomised controlled trial [ACTRN12610000607077]

Roberts, CL, Kolling Institute, University of Sydney, Australia
Rickard, KR, Kolling Institute, University of Sydney, Australia
Algert, CS, Kolling Institute, University of Sydney, Australia
Morris, JM, Kolling Institute, University of Sydney, Australia

Prevention of preterm birth remains one of the most important challenges in maternity care. We are currently undertaking a randomised trial in New South Wales, Australia that utilises routinely collected data to determine trial outcomes and some of the baseline characteristics.

The aim of this presentation is to report how women recruited into the trial flowed through into follow-up. These initial results are based on the single hospital recruiting in 2011, the first year of trial operation. The trial aim is to evaluate whether treating women with asymptomatic vaginal candidiasis early in pregnancy is effective in preventing spontaneous preterm birth.

This prospective, randomised, open-label, blinded-endpoint (PROBE) study utilises a Candida testing protocol that is easily incorporated into usual antenatal care; a simple, well accepted, treatment intervention; and assessment of outcomes from validated, routinely-collected, computerised databases. The intervention is a 6-day course of clotrimazole vaginal pessaries (100mg) and the primary outcome is spontaneous preterm birth <37 weeks gestation.

Pregnant women presenting for antenatal care <20 weeks gestation with singleton pregnancies are eligible for recruitment. Women who agree to participate self-collect a vaginal swab and those who are culture-positive but asymptomatic for candidiasis are randomised (central, telephone) to open-label treatment or usual care (screening result is not revealed, no treatment, routine antenatal care). Outcomes are obtained from hospital discharge and obstetric databases. A sample size of 3,208 women with asymptomatic candidiasis (1,604 per arm) is required to detect a 40% reduction in the spontaneous preterm from 5.0% in the control group to 3.0% in women treated with clotrimazole (significance 0.05, power 0.8). Analyses will be by intention to treat.

In the first year of recruitment at the first participating hospital, 795 women (77% of those approached) were recruited. Of these 154 (19%) had asymptomatic Candida colonisation and were randomised. The subsequent birth outcome information was readily available for the 745 (94%) recruited women who delivered at that same hospital. However 50 (6%) women gave birth elsewhere and their outcome information will ultimately need to be obtained through statewide record linkage. Women both with and without readily available outcome data were similar for baseline characteristics including age, parity and gestation at recruitment. Among the first year recruitments with outcome data, the preterm birth rate (4.1%) was lower than the estimate of 5% based on all-hospital data, which could have implications for the study's power to find an effect.

Corresponding author email: christine.roberts@sydney.edu.au

Navigating the challenges of record linkage in a multi-centre research study across two health boards

Strachan, F E, University of Edinburgh
Shah, A S, University of Edinburgh
Mills, N M, University of Edinburgh

Introduction

High-sensitivity troponin assays can now detect troponin in the majority of healthy persons (Apple, 2013) with this development directly leading to new guidelines for the definition of myocardial infarction (Thygesen, 2012). Lowering the diagnostic threshold with these assays could improve diagnostic accuracy and clinical outcome in patients with suspected acute coronary syndrome. However, troponin is elevated in many conditions and inappropriate diagnosis of myocardial infarction may be detrimental to patient care.

Using routine data and electronic patient records we aim to establish whether the introduction of a high-sensitivity troponin assay into routine clinical practice is beneficial or harmful for patient care.

Methods

In a cluster, randomised controlled study across 10 hospitals in Scotland, in NHS Greater Glasgow and Clyde and NHS Lothian, we will identify patients with suspected acute coronary syndrome who have troponin measured as part of their routine assessment. Following a 6-month validation phase, each site will be randomised to early or to late implementation of a high-sensitivity troponin I assay.

Clinical information will be collected to characterise the patient’s initial episode and to monitor clinical outcomes. Data will be linked from a number of electronic data sources, including the TrakCare electronic patient record application; laboratory reporting systems; the MUSE ECG archive; ECHOPACS echocardiography database; TOMCAT angiography database; ISD SMR and prescribing records.

We have ethical approval to link patient data without individual consent (Reference 12/SS/0115) and Caldicott Guardian approval for data management processes within NHS Lothian. Further approvals will be sought from the Caldicott Guardian for NHS Greater Glasgow & Clyde and PAC for linkage of ISD data sets.

Discussion

Routine electronic patient information and data sources provide important opportunities to improve patient care through evaluation of major changes in clinical practice. However, there are challenges in linking datasets from different platforms and across different sites. The variability in the uptake and implementation of systems within and between health boards adds to the complexity of data management and analysis processes.

Suspected acute coronary syndrome is the most common reason for unplanned admission to hospital and despite improvements in the management of myocardial infarction one-year mortality remains high. Through record linkage this study will provide a critical evaluation of the modern definition of myocardial infarction and will lead to better care for patients with suspected acute coronary syndrome across Scotland.

References

Thygesen K, Alpert JS, Jaffe AS, et al. Third Universal Definition of Myocardial Infarction. European Heart Journal (2012) 33, 2551-2567

Apple F, Ler R, Murakami MA. Determination of 19 Cardiac Troponin I and T Assay 99th Percentile Values from a Common Presumably Healthy Population. Clinical Chemistry (2013) 58:11

Corresponding author email: f.strachan@ed.ac.uk

Using linked healthcare data to create a randomised controlled trial: The RAPiD Trial (Reducing Antibiotic Prescribing in Dentistry)

Elders, A, University of Aberdeen
Prior, M, University of Aberdeen
Clarkson, J, NHS Education for Scotland
Duncan, E, University of Aberdeen
Elouafkaoui, P, NHS Education for Scotland
Ramsay, C, University of Aberdeen
Young, L, NHS Education for Scotland

Routinely-collected electronic healthcare data can be linked and used in several ways in a randomised controlled trial (RCT). We report the design of an RCT of dental prescribing behaviour which uses linked data in five aspects of the trial:

1) to identify participants;
2) to apply inclusion/exclusion criteria;
3) to carry out stratified randomisation;
4) to generate the trial intervention;
5) to analyse trial outcomes;

Antibiotic prescribing in dentistry accounts for 9% of total antibiotic prescriptions in primary care in Scotland. The Scottish Dental Clinical Effectiveness Programme (SDCEP) published guidance In April 2008 (updated in August 2011) for drug prescribing in dentistry. The SDCEP guidance includes information to assist dentists to make evidence-based decisions about antibiotic prescribing, but wide variation in prescribing persists and the overall use of antibiotics is increasing. The aim of the 12 month three-arm RCT is to compare the effectiveness of enhanced audit and feedback strategies for the translation into practice of the SDCEP antibiotic prescribing guidance across Scotland. The trial is being conducted as part of the TRiaDS (Translation Research in a Dental Setting) programme.

Information relating to General Dental Practitioners (GDPs) and dental practices were obtained from routinely held workforce data. Permission was granted to use data from the Management Information and Dental Accounting System (MIDAS) database and the Prescribing Information System for Scotland (PRISMS) database. We linked these two datasets to produce individualised feedback on antibiotic prescribing for all GDPs in all fourteen Scottish Health Boards. These data were subsequently linked to the dental workforce data so that comparable feedback could be posted to all currently practicing GDPs

Eligibility was determined on contract status and a minimum level of recent treatment provision. All eligible dental practices in Scotland were simultaneously randomised at baseline either to current audit practice or to one of two audit and feedback interventions. Randomisation was stratified by single-handed/multi-handed practices. All GDPs working in a practice allocated to an intervention group received graphical representations of their antibiotic prescribing rate.
Intervention practices were further randomised using a factorial design to receive feedback either with or without a health board comparator and with or without a supplementary text based intervention. The initial feedback report contained retrospective prescribing data taken from the previous 14 months. GDPs in a randomly assigned 50% of the intervention practices receive two reports (at 0 and 6 months); those in the other practices receive three reports (0, 6 and 9 months). The primary outcome is the total antibiotic prescribing rate per 100 courses of treatment over the year following the delivery of the baseline intervention.

In describing the design of this study, we demonstrate that linked administrative datasets have the potential to be used efficiently and effectively across all stages of a randomised trial. We also discuss the various challenges and limitations that such an approach presents.

Corresponding author email: andrew.elders@abdn.ac.uk

The use of routinely collected health data in clinical trials: an example from the SELENIB bladder cancer trial

Evison, F, Health Informatics Department, University Hospitals Birmingham NHS Foundation Trust, United Kingdom
Begaj, I, Health Informatics Department, University Hospitals Birmingham NHS Foundation Trust, United Kingdom
Devall, A, School of Cancer Sciences, University of Birmingham, United Kingdom
Pirrie, S, Cancer Research Clinical Trials Unit, School of Cancer Sciences, University of Birmingham, United Kingdom
Patel, P, Department of Urology, University Hospitals Birmingham NHS Foundation Trust, United Kingdom
James, N, School of Cancer Sciences, University of Birmingham, United Kingdom
Bryan, R, School of Cancer Sciences, University of Birmingham, United Kingdom

Bladder cancer is the seventh most common cancer in the UK, with current estimates suggesting that one in 40 men and one in 107 women are likely to develop bladder cancer over their lifetime. Cancer Research UK (CRUK) estimated that in 2007 the ten year survival of patients with bladder cancer was 48.9%.

From 2007-11 researchers at the University of Birmingham (UoB), Queen Elizabeth Hospital Birmingham (UHB NHSFT) and members of the Birmingham CRUK Centre recruited participants to a 2x2 placebo-controlled randomised trial of selenium and vitamin E for preventing recurrence and progression in patients with non-muscle-invasive bladder cancer (“SELENIB”), enrolling patients from the West Midlands.

After the trial closed to recruitment, it was decided that further information about the patient pathways would be beneficial to supplement the data already collected. To achieve this, it was deemed that robust data could be obtained from routinely collected datasets (i.e. Hospital Episode Statistics (HES), ONS Mortality). Moreover, the HES dataset could be used to corroborate the prospectively-collected trial data and potentially reduce the cost/errors of manual collection. To perform the linkage of trial data to HES, a number of issues regarding governance, ethics and process were raised.

Following the terms of the trial patient consent form, all analysis of medical records would be done by members of the SELENIB team. Researchers at QEB had access to HES data, but at an anonymised level with access to local hospital identification numbers.

We describe how we overcame concerns to ensure that patient confidentiality was not breached and that all governance issues were safely dealt with. Other issues that we considered included the point at which the HES analysis could be safely conducted without breaching trial protocol. We also describe issues surrounding the actual linkage of HES data to that obtained in the trial, and provide recommendations as to what data might be collected from trials in the future to try and improve matching, and how patient consent forms may be amended to ensure that informed consent is obtained for HES data matching.

HES could potentially be a rich data source for use within clinical trials. However, there are a number of data items not currently collected that will be mandatory for HES to be considered as a data source to collect key trial information. Not least of which would be the legal requirement of coding procedures and diagnoses in outpatient data.

Corresponding author email: Felicity.Evison@uhb.nhs.uk

Understanding the impact of fertility history on outcomes in mid-life in Scotland, a longitudinal approach using the Scottish Longitudinal Study (SLS)

Williamson, LEP, Longitudinal Studies Centre - Scotland (LSCS), University of St Andrews
Dibben, C, Longitudinal Studies Centre - Scotland (LSCS), University of St Andrews

This study is part of the research programme involving data linkage within the Scottish Health Informatics Programme (SHIP). The research draws on and extends work on reproductive histories and life outcomes. Previous studies have shown that the number of children (parity) can be linked to specific health outcomes in mid and later life for women (references can be provided). We aim to extend this research specifically for Scotland based on Scottish data, namely the Scottish Longitudinal Study (SLS) linked to health data from the NHS Scottish Morbidity Record (SMR) datasets, including the maternity dataset SMR02 (as parity is only recorded for married women at birth registration in Scotland).

The aim of this SHIP project, involving data linkage and health outcomes, is to gain a full understanding of the impact of both fertility histories and childlessness on health outcomes and mortality. In addition, we plan to compare findings with previous research where applicable. This research is only for specific female SLS birth cohorts, as it is acknowledged that we are not able to follow-up all SLS members or SLS members to old ages since the SMR02 is only available from 1975. Nevertheless, the SLS allows follow-up of the specific SLS birth cohorts from the 1991 Census until 2009 (the most recent year death data is available linked to the SLS). From preliminary modelling, in line with previous research, we find high birth parity to be an important factor in relation to mortality.

Corresponding author email: lee.williamson@st-andrews.ac.uk

Social Housing and Health in Manitoba: A first look

Smith, M, Manitoba Centre for Health Policy, University of Manitoba
Finalyson, G, Manitoba Centre for Health Policy, University of Manitoba
Martins, P, Manitoba Centre for Health Policy, University of Manitoba
Dunn, J, University of Toronto
Soodeen, RA, Manitoba Centre for Health Policy, University of Manitoba
Prior, H, Manitoba Centre for Health Policy, University of Manitoba
Taylor, C, Manitoba Centre for Health Policy, University of Manitoba
Burchill, C, Manitoba Centre for Health Policy, University of Manitoba
Guenette, W, Manitoba Centre for Health Policy, University of Manitoba
Hinds, W, Manitoba Centre for Health Policy, University of Manitoba

Fourteen years of social housing data (1995-2008) were acquired from the provincial government and linked to the population health research data repository at the Manitoba Centre for Health Policy. This allowed an opportunity to compare those living in social housing to the rest of the population in Manitoba on a number of health and social indicators.

Comparisons were made on 19 indicators of morbidity, mortality, health care utilization and social development. Logistic regression models were developed to control for variations in age, sex, region of residence, presence of comorbidities, income, and neighborhood level SES, and other confounding factors.

Fifty percent of the social housing population is under the age of 20, 75% are female and 50% of applicants receive income assistance. As expected for such a low income group there are significant differences on most health status measures when compared to individuals not living in social housing. However, after controlling for income most differences between the two groups disappeared indicating that there is no independent effect of social housing. The exceptions were total respiratory morbidity, mammography and high school completion. And in two other cases, cervical cancer screening and complete immunization by age two, the modeled rates were actually better for individuals in social housing than for a comparable group of low-income individuals not in social housing. High school completion rates showed a very significant interaction with neighborhood level SES. The policy implications of this research are discussed.

Corresponding author email: mark_smith@cpe.umanitoba.ca

Ethnic differences in upper gastrointestinal disease in Scotland

Ward, H, NATIONAL SERVICES SCOTLAND
Brin, GI, University of Edinburgh
Bansal, N, University of Edinburgh
Bhopal, R, University of Edinburgh
Bhala, N, University of Oxford

Background and objectives

There is a paucity of data assessing ethnic variations in upper gastrointestinal (GI) disease: we studied the incidence of upper GI diseases using adequate measure of ethnicity in Scotland, made available by record linkage to the Scottish Census 2001.

Methods

Using the Scottish health and ethnicity linkage study (SHELS), linking NHS hospital admissions and mortality to the Scottish census 2001, we explored ethnic differences in incidence (2001-2010) of specific upper GI diseases (peptic ulcer disease, oesophagitis, gastritis, gallstones and pancreatitis) in Scotland. Risk ratios (RR) were calculated using Poisson regression with robust variance and multiplied by 100, by gender, adjusted for age and subsequently country of birth (COB) and socio-economic status using Scottish Index of Multiple deprivation (SIMD) available in the Census. The White Scottish population was the reference population (100).

Results

The total numbers of first events within the 9 years period of interest (over almost 29 million of Person-Year (PY) at risk) was 44,612 for peptic ulcer, 102,706 for oesophagitis, 141,235 for gastritis, 87,556 for gallstones and 17,177 for pancreatitis. Looking at risk ratios for all specific upper GI diseases and compared to respectively White Scottish men and women, other White British and other White had a lower risk of upper GI diseases. White Irish had an increased risk of upper GI diseases which did not remain after adjustment for COB and SIMD.

There were consistent ethnic variations in non-White minority ethnic group even after adjustment for COB and SIMD. Peptic ulcer disease risk was increased in Chinese men (RR[CI]=171.0 [131.4;222.5]) and other South Asian men (146.0 [106.8;199.6]) and women (160.5 [101.8;252.9]). Pakistani and Bangladeshi had an increased risk of esophagitis whereas Chinese had a lower risk (65.4 [50.9; 83.9] for men, 69.3 [54.5; 88.0] for women). South Asian had an increased risk of gastritis whereas it was lower for men of African origin (62.5 [40.4; 96.6]). Gallstones was more incident in Chinese men (141.6 [105.9; 189.2]) and Pakistani women (128.6 [112.1; 147.6]), the latter also had an increased risk of pancreatitis (147.1 [108.8; 198.8]).

Conclusions

This unique linked data allowing the comparison of specific upper GI diseases incidences between ethnic groups has shown major differences. These findings are unprecedented and show the value of record linkage in ethnicity and health research. Further linkage would be required to explore risk factors and understand these differences in order promote health equality.

Corresponding author email: hester.ward@nhs.net

The Effect of Ethnicity on Outcomes Following Emergency Surgery: A national cohort study

Vohra, R, Academic Department of Surgery, University Hospitals Birmingham NHS Foundation Trust, UK
Evison, F, Health Informatics Department, University Hospitals Birmingham NHS Foundation Trust, United Kingdom
Begaj, I, Health Informatics Department, University Hospitals Birmingham NHS Foundation Trust, United Kingdom
Patel, P, Department of Urology, University Hospitals Birmingham NHS Foundation Trust, United Kingdom
Pinkney, T, Academic Department of Surgery, University Hospitals Birmingham NHS Foundation Trust, United Kingdom

According to the 2011 Consensus, 13% of UK residents are born outside the UK and 8% are unable to speak English. In some areas, especially around London, >70% of the population are predominately non-white. Rapid assessment and treatment is essential in patients admitted with acute surgical pathology. Delays can produce profound and prolonged effects on outcomes. There are clearly language barriers when treating patients from an ethnic background. In addition, there are cultural barriers such as omissions of sensitive elements in the history, adequate expose patients for examinations, appropriate chaperones, etc. These are difficult to quantify individually, but may have a cumulative effect in delaying diagnosis and treatment.

Using Hospital Episode Statistics (HES) we identified all adult patients who were admitted to hospital as an emergency and underwent a surgical procedure between April 2007 and March 2012. We identified these patients using procedure codes from the Agency for Healthcare Research and Quality (AHRQ). Where a valid ethnicity was not recorded for the index admission, other in hospital stay and outpatient spells were investigated. The index admission was linked to ONS mortality records. In addition, we investigated if there were regional variations in emergency surgical admission, mortality and geographical location within the UK as recorded in the 2011 census.

The number of patients with invalid ethnicity recorded, dropped from 7% to 4% during this time period. Initial crude results suggest that these patients with unknown ethnicity codes are more likely to die following emergency surgery than those whose ethnicity is recorded.

There are many factors which could influence the outcomes of a patient following emergency surgery. We will further explain some of the methods we used to account for such differences to allow a more direct comparison between ethnic groups. We then present the results of these analyses.

Corresponding author email: Felicity.Evison@uhb.nhs.uk

Using linked data to analyse rates if intervention in childbirth in England by mothers' countries of birth

Macfarlane, AJ, City University London
Dattani, N, City University London
Datta-Nemdharry, P, City University London

Background
There is concern that women from some migrant groups may experience unjustifiably high rates of obstetric intervention but it is difficult to investigate this using routinely collected data for England and Wales. Patients' ethnicity but not their country of birth is recorded on NHS records, including those recording data about care at delivery.

Aim
To analyse data about care at delivery according to parents origins

Methods
Using methods described previously, data from birth registration, which includes parents' countries of birth we linked to delivery records from the Maternity Hospital Episode Statistics for England to enable data about care at delivery to be tabulated by parents' countries of birth.

Results
Variations in caesarean section rates for individual mothers' countries of birth were wider than differences between ethnic groups. Rates of caesarean section by mothers' country of birth varied both within and between regions of Africa, with some countries having very high rate. Analyses for other continents will also be presented.

Discussion
Record linkage added value to existing data by enabling analyses of care given to be analysed by mothers' countries of birth as wall as by ethnicity and to identify groups with high levels of intervention.

Corresponding author email: A.J.Macfarlane@city.ac.uk

Caesarean section rates - using routinely collected data to examine inter-hospital variation

Lee, YY, Kolling Institute, University of Sydney, Australia
Patterson, J, Kolling Institute, University of Sydney, Australia
Ford, JB, Kolling Institute, University of Sydney, Australia
Simpson, J, School of Public Health, University of Sydney, Australia
Roberts, CL, Kolling Institute, University of Sydney, Australia

Background: International concern over rising caesarean section rates has focused attention on initiatives to reverse this trend. We assessed variation in caesarean rates among hospitals to identify potential targets for reducing practice variation associated with high rates.

Methods: This is a population-based of 183,310 maternities in 81 hospitals in New South Wales, Australia, 2009-2010. Data were obtained from longitudinally-linked birth records for each woman. Deliveries were categorised into ten risk-based, mutually exclusive groups according to the Robson classification. These ten groups are inclusive of all caesareans, and are based on parity, plurality, labour onset, previous caesarean, fetal presentation and gestation. Multilevel logistic regression was used to examine variation in hospital caesarean rates within Robson groups, adjusted for differences in maternal age, ethnicity, smoking, diabetes, hypertension and type of maternity care. The 20th centile ("best practice rate") of the risk-adjusted rates was used to quantify the potential impact on the overall caesarean rate of reducing practice variation.

Results: The overall caesarean rate was 30.9%, ranging from 11.8% to 47.4% across hospitals. Previous caesarean (36.4% of all caesareans) and nulliparous term births (elective 23.4% and spontaneous 11.1%) were the greatest contributors to the overall rate. After adjustment, marked unexplained variation in hospital caesarean rates persisted for: nulliparous women at term, previous caesarean, multiple pregnancies and preterm births. For example, the caesarean rate among nullipara with spontaneous labour at term ranged from 8.1% to 23.5%, and amongst women with a previous caesarean section from 67.0% to 91.9%. If variation in practice was reduced by achieving the "best practice" rate for these risk-based groups, this would lower the overall caesarean rate by 3.1%.

Conclusions: Reducing unwarranted practice variation is important where it influences health outcomes, health care costs, and provision of appropriate and patient-focused care. Understanding the extent of hospital heterogeneity in caesarean practice and implementing evidence-based practices may result in improved maternity care. We have identified five risk-based groups as priority targets for reducing practice variation and caesarean births.

Corresponding author email: christine.roberts@sydney.edu.au

Inherited Risk of Pre-eclampsia: using two approaches for analysis

Bhattacharya, S, Lecturer, Obstetric Epidemiology, University of Aberdeen
Raja, EA, Medical Statistics Team, University of Aberdeen
Campbell, DM, Emeritus Reader, Obstetrics and Gynaecology, University of Aberdeen

Background: Several previous research reports have suggested a genetic predisposition to pre-eclampsia but none have demonstrated the effect separately in nulliparous and parous women in the context of other risk or protective factors.

Objective: To assess the magnitude of genetic predisposition to pre-eclampsia with reference to other risk factors in nulliparous and parous women.

Material and Method: The Aberdeen Maternity and Neonatal Databank records all pregnancy and delivery details occurring in Aberdeen, Scotland since 1950. It has now become possible to link pregnancy records of mothers and grandmothers to those of the daughters. Using a nested case control design within this intergenerational cohort, statistical modelling was done with known risk/protective factors for pre-eclampsia, separately for nulliparous and parous women. Conditional logistic regression was used to compare characteristics between parous pre-eclamptics and year and parity matched normotensive controls.
In a separate analysis, including all parous women, we used a multilevel approach based on Generalised Estimating Equation (GEE) and specified the link function as binomial. We assumed a working exchangeable correlation of having preeclampsia within a daughter in her pregnancies. Odds ratios (OR) and 95% confidence intervals (CI) were estimated through GEE with the use of robust standard errors.

Results: There were 34,970 mother-daughter pairs. Of the daughters, there were 1,248 nulliparous and 448 parous pre-eclamptics. For nulliparous women, the risk factors remaining in the stepwise model were mother's history of pre-eclampsia {O.R. 2.13 (95% C.I. 1.57, 2.89)}, booking BMI >30Kg/m2 {O.R. 2.06 (95% C.I. 1.68, 2.52)}, age, gestation period, and booking diastolic blood pressure. Smoking “ 10 cigarettes a day was protective against pre-eclampsia {O.R. 0.52 (95% C.I. 0.44, 0.62)}. For multiparae, the risk factors included pre-eclampsia in the initial pregnancy {O.R. 8.80 (95% C.I. 1.54, 50.23)}, advanced age at delivery {O.R. 3.09 (95% C.I. 1.69, 5.66)} and BMI >30 Kg/m2 {O.R. 2.61 (95% C.I. 1.62, 4.20)}. Smoking 10 or more cigarettes per day was protective {O.R. 0.57 (95% C.I. 0.35, 0.94). A history of maternal pre-eclampsia was not independently associated with an increased risk of development of pre-eclampsia in the multiparae after adjusting for other covariates. Results were similar using the GEE approach.

Conclusion: In nulliparous women, a history of maternal pre-eclampsia was associated with more than doubling of risk of pre-eclampsia. In multiparae, this association was not observed, although a history of pre-eclampsia in a previous pregnancy was strongly associated with increased risk, suggesting genetic susceptibility.

Corresponding author email: sohinee.bhattacharya@abdn.ac.uk

Recurrence Risk of Obstetric Anal Sphincter Injury (OASI) During Childbirth in New South Wales, Australia.

Ampt, AJ, Kolling Institute, University of Sydney, Australia

Background:
Obstetric anal sphincter injury (OASI), in which major tearing of the perineum during childbirth extends into the anal sphincter, has short and long term consequences for women. For a woman having a subsequent birth, major damage can occur again, resulting in an OASI recurrence. A history of an OASI thus raises clinical questions regarding mode of delivery and potential risk for subsequent births. Longitudinal linkage of birth and hospital discharge data allows for investigation of recurrence risk. To date, studies investigating factors associated with recurrence have been unable to assess the role of first birth factors.

Aim:
To determine the OASI recurrence rates at the second birth for women who have previously sustained an OASI at their first birth; to identify the modes of delivery for the second birth; and to identify significant predictive factors for recurrence from both the first and second births.

Methods:
Data were obtained from two linked NSW population-based collections - the Perinatal Data Collection (a statutory data collection for all births “20 weeks gestation or “400g birthweight), and the Admitted Patients Data Collection (a census of discharge data from hospital records). These were lined using probabilistic methods. The study population consisted of women with at least a first and second consecutive birth during 2001-2010, where the second was term (37- 41 weeks gestation) and vertex. Predictive factors for recurrence were determined using multivariable logistic regression.

Results:
Among women with a first birth OASI, 22% had a caesarean section for their second birth. Among second vaginal births, the OASI recurrence rate was 5.5%, with risk factors from the first birth identified as episiotomy with an adjusted odds ratio (aOR) of 1.47 and birthweight < 3kg (aOR 1.62); and from the second birth forceps delivery without episiotomy (aOR 11.42), regional analgesia (aOR 1.92) and birthweight > 4kg (aOR 2.56). Protective factors from the first birth included regional analgesia (aOR 0.69), and from the second, oxytocin (aOR 0.61).

Conclusion:
An OASI recurs for one in twenty women who have a second vaginal birth. Knowledge of the risk factors for an OASI recurrence can provide information to clinicians and women to help inform decisions around mode of subsequent delivery. Longitudinal record linkage makes possible the assessment of the previous birth impact on the subsequent birth.

Corresponding author email: amanda.ampt@sydney.edu.au

The benefits of using linked primary and secondary care data in developing a population based co-morbidity score

Crooks, C J, University of Nottingham
West, J, University of Nottingham
Card, T R, University of Nottingham

Introduction
Hospital derived co-morbidity scores such as the Charlson index are frequently used to adjust for co-morbidity in population based studies. We assessed whether using the combined information from linked primary and secondary care records provided an opportunity to develop an improved population based score of co-morbidity.

Methods
We used the linked Hospital Episodes Statistics, Clinical Practice Research Datalink, and Office of National Statistics death register. We selected all people older than 20 years at the start of 2005 and followed them up until 2010. A random 50% sample was used for the development of the score, and then internal validation was performed in the remaining 50% of the cohort. A Bayesian hierarchical model was used to select codes and groupings associated with a reduced survival. The hierarchy was based on the Read code system, into which relevant ICD 10 codes had been mapped. The diagnoses coded for each individual were then used in a Cox proportional hazard model to construct a co-morbidity score predicting 5 year survival from the start of 2005. Weightings were derived based on adjusted coefficients from the model that were greater than one. Different weightings were used when the coefficients differed between hospital and primary care. Discrimination was assessed using Harrell's C statistic and comparisons made with the Charlson index over time, age and consultation rate.

Results
659,706 people were followed up from the 1st January 2005. 93 groupings of codes were derived from the Bayesian hierarchy, and 40 of these had an adjusted weighting of greater than one in the Cox proportional hazards model. 37 of these groupings had a different weighting dependent on whether they were coded from hospital or primary care data. The C statistic reduced over the follow up period from 0.878 in the first year to 0.853 over all 5 years. The Charlson index had a C statistic of 0.847 over the 5 years of follow up. When we stratified our linked score by consultation rate the association with mortality remained consistent, although the discrimination was reduced among patients with higher consultation rates.

Conclusions
We developed a co-morbidity score in linked population based primary and secondary care data. However our attempt to improve on existing scores through an increase in complexity resulted in only small improvements in discrimination. Nevertheless we have demonstrated that such scores can usefully predict mortality far beyond the 1 year for which they were initially developed.

Corresponding author email: colin.crooks@nottingham.ac.uk

Charlson scores derived from administrative data and case-note review compared favourably in a population-based cohort

Johnston, M C, NHS Grampian / University of Aberdeen
Marks, A, University of Aberdeen
Crilly, M, NHS Grampian / University of Aberdeen
Prescott, G, University of Aberdeen
Robertson, L, University of Aberdeen
Black, C, NHS Grampian / University of Aberdeen

Background

Comorbidity describes the burden of disease coexisting with a particular disease of interest. It can affect the course and outcome of disease or illness and is an important confounding factor as well as being predictive of outcomes. Consequently, the accurate assessment of its impact is important clinically, for example for case management and health-care planning, as well as in research. The Charlson index is a widely used measure of comorbidity and was originally developed using case-note review data, often considered to be the gold-standard method. However, the case-note review process is resource intensive and as a result the Charlson index has been adapted for use with routine administrative datasets. Despite this, there remains uncertainty about the appropriateness of its application to administrative data and no previous study has compared Charlson scores derived from case-note review to those derived from administrative data in a chronic kidney disease population.

Objectives

The objective was to compare Charlson index scores calculated using administrative data to those calculated using case-note review in relation to all-cause mortality and initiation of renal replacement therapy (RRT) in a Scottish population based chronic kidney disease cohort.

Methods

Modified Charlson scores were calculated for each individual from both case-note review and administrative data in the Grampian Laboratory Outcomes Morbidity and Mortality Study (GLOMMS-1) cohort. Agreement between scores was assessed using the weighted kappa. The association with outcomes was assessed using Poisson regression and the performance of each was compared using net reclassification improvement.

Results

Of 3,382 individuals, median age 78.5 years, 56% female, there was moderate agreement between scores derived from the two data sources (weighted kappa 0.41). Both scores were associated with mortality independent of a number of confounding factors. Administrative data Charlson scores were superior to case-note review scores at classifying the risk of death using net reclassification improvement. Neither score was associated with commencing RRT.

Conclusions

Despite only moderate agreement, Charlson scores from both data sources were associated with mortality. Neither was associated with commencing RRT. Administrative data compared favourably, and may be superior, to case-note review when used in the Charlson index to predict mortality.

Corresponding author email: marjorie.johnston@nhs.net

Multimorbidity: impact on health systems and their quality of care

Yu, N, Division of Population Health Sciences, University of Dundee
Guthrie, B, Division of Population Health Sciences, University of Dundee
Mercer, S, Institute of Health and Wellbeing, University of Glasgow

Background

The management of people with multimorbidity poses increasing challenges for health care systems, which remain largely configured for the care delivery of single-disease. Quality of care is generally better for people with multiple conditions compared to those with only one, possibly because higher consultation rates provide more opportunities to deliver care. However, most studies have not distinguished different types of co-existing morbidity.

Aim

The aim of this study was to examine quality of diabetes care by co-morbidity count and by count of co-morbid physical and mental health conditions.

Method

Data on the presence of 40 morbidities and quality of diabetes care was extracted for 58,593 people with type 2 diabetes registered with 314 general practices in Scotland. Two “all or none” composite measures were created, defined as whether or not a patient received all four specified processes (HBA1c, BP, cholesterol and smoking recorded in the previous 12 months) or achieved all four specified intermediate outcomes (HBA1c⩽7.4%, BP ⩽140/80mmHg, cholesterol⩽5.0mmol/l, not smoking). Co-morbidity was defined in two ways: first, as a simple co-morbidity count (the approach taken by most previous studies); second, as a separate count of the number of “concordant” physical conditions (eg angina), the number of “discordant” physical conditions (eg COPD), and the number of mental health conditions. Associations were examined with logistic regression, adjusting for age, sex and social deprivation.

Results

65.7% of patients received all four processes in the previous year, and 13.1% achieved all four intermediate outcome targets. Increasing co-morbidity was associated with better quality of care in a stepwise manner (for 5 or more co-morbidities vs none, process composite adjusted OR 1.34, 95%CI 1.18-1.36; outcome composite adjusted OR 1.35, 95%CI 1.21-1.50). Concordant physical comorbidity was also associated with better process and outcome quality, but there was only a weak positive association with the number of discordant physical conditions, and mental health co-morbidity was associated with worse quality of care (for 2 or more mental health co-morbidities compared to none, process composite OR 0.87, 95%CI 0.82-0,93; outcome composite OR 0.82, 95%CI 0.75-0.89).

Conclusions

Despite high quality on individual measures, reliable delivery of processes and intermediate outcomes in type 2 diabetes is relatively poor. As previous studies have shown, co-morbidity is associated with better quality of care. However, this broad association conceals differences between different types of co-morbidity. “concordant” co-morbidity is associated with better quality of care, but associations with “discordant” co-morbidity are weak, and mental health co-morbidity is associated with worse quality of care. Improving quality for people with physical-mental health co-morbidity is a key challenge for health services.

Corresponding author email: n.yu@dundee.ac.uk

Using hospital data to identify comorbidity and multimorbidity in Australia

Lujic, S, University of Western Sydney
Jorm, L, University of Western Sydney; The Sax Institute
Randall, D, University of Western Sydney

Background

The degree to which administrative data include complete and accurate comorbidity and multimorbidity information, commonly used in risk adjustment, is relatively unknown. Our study investigated the level of agreement between self-reported and administrative data collections, and identified patient and facility-level factors that impact on the agreement levels.

Methods

The 45 and Up Study cohort includes 266,848 people aged 45 years and over from across New South Wales (NSW), Australia's most populous state. Self-reported morbidity data from participants in the 45 and Up Study was linked with hospital data from the NSW Admitted Patient Data Collection (APDC) for the 365 days prior to completion of the Study questionnaire. Six morbidity codes and 15 two-way comorbidity combinations were examined, both on the most recent hospital record ("index") and up to 1 year prior to the "index" record ("lookback"). The overall agreement measure was assessed using Cohen's Kappa (κ). Multivariate logistic regression analyses were conducted to indentify individual- and facility-level factors associated with agreement between the two data sources.

Results

A total of 32,832 study participants were admitted to 314 hospitals up to a year prior to filling out the 45 and Up Study baseline questionnaire. 72% of participants reported at least one, and 36% reported at least two morbidities. Index record agreement was found to be good for diabetes (κ =0.79), moderate for smoking (κ =0.59), fair for heart disease, hypertension and stroke (κ =0.24 - 0.40), and poor for obesity (κ =0.09). Sensitivities ranged from 74% for diabetes to 7% for obesity, indicating that a large number of individuals with self-reported morbidities did not have a record of such morbidity on their index hospitalisation. Incorporation of a lookback period increased morbidity ascertainment for all conditions, particularly obesity, but agreement was still poor (κ =0.13). Patient factors associated with better agreement between self-reported and hospital records included younger age, male sex, emergency rather than planned admission, surgical admission and lower income (for obesity only). The percentage of unexplained variation due to the hospital of admission varied between 7% (diabetes) and 21% (obesity). Hospital-level factors associated with better agreement included public and major city hospitals.

Conclusions

The completeness of capture of common comorbid conditions in routine hospital data is highly variable, and for some conditions, very low. Data linkage, with longer lookback periods, can help enhance ascertainment. However, variation in completeness of capture according to hospital has the potential to introduce bias into risk-adjusted comparisons of hospital performance.

Corresponding author email: s.lujic@uws.edu.au

Burn injury and cancer risk: A record-linkage study using data from Western Australia and Scotland.

Duke, JM, Burn Injury Research Unit, University of Western Australia, Australia
Boyd, J, Curtin University, Australia
Bauer, J, Curtin University, Australia
Rea, S, Burn Service of Western Australia, Royal Perth Hospital, Princess Margaret Hospital for Children, Western Australia
Wood, F, Burn Service of Western Australia, Royal Perth Hospital, Princess Margaret Hospital for Children, Western Australia

Background: While burns predominantly affect the skin, burns can have significant effects on the whole body, and are associated with depressed immune functioning and prolonged periods of metabolic changes. These metabolic changes also promote the division of somatic, non-lymphoid cells which subsequently increase the potential for malignant transformation. A diagnosis of hepatocellular carcinoma in a young male burn patient, by burn clinicians of this research group, prompted investigation of the risk of cancer in burn survivors. Preliminary research of cancer incidence post-burn injury cancer incidence using linked Western Australia data identified an apparent increased risk for females. To enable investigation of more detailed gender and site-specific cancer incidence, data were sought from both the Western Australia Data Linkage System (WADLS) and ISD Scotland.

Aim: To assess if there is an increased risk of cancer amongst survivors of burn injury.
Methods: Linked hospital morbidity, cancer and death data were obtained from the Western Australia Data Linkage System (WADLS) and ISD Scotland for all persons hospitalised for a first burn injury from 1983 to 2008, respectively, in Western Australia (WA) and Scotland. The cancer incidence of those hospitalised for burn was compared to the general population of WA and Scotland, respectively and Standardised Incidence Ratios and 95% confidence intervals (SIR; 95%CI) were calculated.

Results: From 1983-2008, there were 23,450 burn patients with 759 post-burn incident cancer notifications in WA, and in Scotland, 37,506 burn patients with 3,372 post-burn cancer notifications, included in the analyses. Preliminary analyses of gender and site-specific cancers have identified consistency of SIR results for cancers of the buccal cavity, liver, oesophagus and respiratory tract. Results of particular initial interest include the significant increase in hepatic cancer risk (combined gender) in both the WA (SIR, 95%CI: 2.6, 1.6-4.0) and Scottish (SIR, 95%CI: 1.7, 1.2-2.5) burn patient data, with females having a higher risk than males (WA SIR (95%CI): 4.7 (2.0-11.4) vs. 2.2 (1.3-3.7); Scotland SIR (95%CI): 1.9 (1.1-3.7) vs. 1.7 (1.1-2.5)). Analyses of gender and site-specific SIRs will be completed by July 2013.

Conclusions: Preliminary results of gender and site-specific analyses have identified an increased risk of some cancers, with consistent results between WA and Scottish burn patient data. There also appears to be a gender effect in relation to the incidence of some site-specific cancers after burn injury, with female burn patients experiencing an increased incidence of cancer. Further exploration of the association of burns and cancer is required to reach a definitive conclusion.

Corresponding author email: janine.duke@uwa.edu.au

Mobile phone use and risk of brain neoplasms and other cancers: prospective study.

Benson, VS, University of Oxford
Pirie, K, University of Oxford
Schüz, J, International Agency for Research on Cancer (IARC)
Reeves, GK, University of Oxford
Beral, V, University of Oxford
Green, J, University of Oxford

Objectives: To use routinely collected cancer registry data and Hospital Episodes Statistics to investigate the relation between mobile phone use and incidence of brain neoplasms and other cancers and conditions in a UK prospective cohort, the Million Women Study.

Methods: Cox regression models were used to estimate adjusted relative risks (RRs) and 95% confidence intervals (CIs).

Full abstract embargoed for publication.

Corresponding author email: vicky.benson@ceu.ox.ac.uk

Agricultural land usage and primary bone cancer: is there a link? Small-area analyses of osteosarcoma and Ewing sarcoma diagnosed in 0-49 year olds in Great Britain, 1985-2009

Blakey, K, Institute of Health and Society, Newcastle University, Newcastle-upon-Tyne, England, United Kingdom
Feltbower, RG, Paediatric Epidemiology Group, University of Leeds, Leeds, England, United Kingdom
Parslow, RC, Paediatric Epidemiology Group, University of Leeds, Leeds, England, United Kingdom
James, PW, Institute of Health and Society, Newcastle University, Newcastle-upon-Tyne, England, United Kingdom
Stiller, C, Childhood Cancer Research Group, Department of Paediatrics, University of Oxford, England, United Kingdom
Norman, P, School of Geography, University of Leeds, Leeds, England, United Kingdom
Garthwaite, D, Food and Environment Research Agency, York, United Kingdom
Hughes, J, Science and Advice for Scottish Agriculture, Edinburgh, United Kingdom
Gerrand, C, Northern Institute for Cancer Research (NICR), Newcastle University and North of England Bone and Soft Tissue Tumour Service, Newcastle-Upon-Tyne, England, United Kingdom
McNally, RJQ, Institute of Health and Society, Newcastle University, Newcastle-upon-Tyne, England, United Kingdom

Objective: This study builds on previous research which found higher incidence of Ewing sarcoma in areas with low population density and high levels of car ownership. Since these factors are characteristic of rural environments the objective was to analyse associations between Ewing sarcoma and osteosarcoma incidence and agricultural land usage.

Methods: All osteosarcoma and Ewing sarcoma cases aged 0-49 years diagnosed in GB during 1985-2009. Pesticide data were a proxy for agricultural land usage and population density an urban/rural indicator. Negative binomial regression was used to examine small area relationships between incidence rates and pesticide levels. The models adjusted for gender, age and deprivation and logarithm of the “at risk” population as an offset.

Results: There were 2562 osteosarcoma cases aged 0-49; 820 aged 0-14; 1262 aged 15-29 and 480 aged 30-49 years. Overall age-standardised incidence rate was 2.84 per million persons per year (95% confidence interval [CI] 2.73 to 2.95). For Ewing sarcoma there were 1711 cases aged 0-49; 670 aged 0-14; 822 aged 15-29 and 219 aged 30-49 years. Overall age-standardised incidence rate was 1.97 per million persons per year (95% CI 1.88 to 2.07). After adjustment for gender, age and deprivation, pesticide usage was not found to have any significant effect on the incidence of osteosarcoma or Ewing sarcoma. For osteosarcoma, the relative risk (RR) for one kilogram per hectare increase in pesticide level = 0.987 (95% CI 0.954 to 1.022) and for Ewing sarcoma RR = 1.011 (95% CI 0.967 to 1.057).

Conclusion: Pesticide usage data and population density were used as proxies for land usage and no association with Ewing sarcoma or osteosarcoma was found. Some other agricultural related factors could explain the higher incidence found in predominantly farmland areas. Further research should investigate other activities such as livestock farming.

Corresponding author email: karen.blakey@newcastle.ac.uk

Policy for home or hospice as the preferred place of death from cancer: Scottish Health and Ethnicity Linkage Study shows challenges across all ethnic groups in Scotland

Sharpe, KH, Information Services Division, NHS National Services Scotland
Brin, G, Public Health Sciences, Edinburgh University Medical School
Bansal, N, Public Health Sciences, Edinburgh University Medical School
Bhopal, RS, Public Health Sciences, Edinburgh University Medical School
Brewster, DH, Public Health Sciences, Edinburgh University Medical School

Background: Place of cancer death varies ethnically and internationally. Palliative care reviews highlight limited ability to demonstrate equal access due to incomplete or unreliable ethnicity data.

Aim: We establish place of cancer death by ethnicity and describe patient characteristics.

Design: We linked census, cancer registry, hospital episode, and mortality data for 117 467 cancer deaths 2001- 2009. With White Scottish population as reference, prevalence ratios (PR) and 95% confidence intervals of death in hospital, home or hospice were adjusted for sex and age at death, calculated by ethnic group.

Results: White Scottish group constituted 91% and non-White ethnic groups combined 0.4% of cancer deaths. South Asian, Chinese and African Origin patients were youngest at diagnosis (63.1, 62.5, 60.7 years) and death (66, 66, and 65.9 years). Hospital death was less likely for White Irish patients (PR= 0.93, 95% C.I. = 0.87; 0.99); home death was more likely (1.15 [1.10; 1.22]). Other White British patients were more likely to die at home (1.07 [1.02; 1.12]); South Asian, African Origin, Chinese and Other ethnic groups combined were less likely to die at home (0.85 [0.74; 0.97]). Generally, affluent Scottish White patients were less likely to die in hospital and more likely to die at home or in a hospice regardless of the socioeconomic indicator used.

Conclusion: Cancer deaths occur most often in hospital (52.3%) for all ethnic groups. Minority ethnic groups combined are less likely to die at home. Regardless of ethnic group, significant work is required to achieve more people dying at or closer to home.

Corresponding author email: katharine.sharpe@nhs.net

An Automated Method for Longitudinally Validating the Presence of Individuals in a Data Set

Thayer, DS, College of Medicine, Swansea University

Introduction: Longitudinal studies using routinely-collected health data raise the problem of verifying individuals' continued presence in the data. Individuals can enter and leave for a variety of reasons, including both life events and the partial coverage of the datasets. An automated, customizable method of determining individuals' presence was developed for the primary care dataset in Swansea University's SAIL Databank.

Methods: The primary care dataset covers only part of Wales, with about 40% of practices participating. The start and end date of the data varies by practice. In addition, individuals can change practices or leave Wales. To address these issues, a two step process was developed. First, the time period for which each practice had data available was calculated by measuring changes in the rate of events recorded over time. Start and end dates were determined by comparing the rate of events to a known good period and using a threshold cut-off. Second, the registration records for each individual were simplified. Anomalies such as short gaps and overlaps were resolved using a set of rules. The result of these two analyses is a set of records indicating start and end dates of available data for each individual. This algorithm was tested on a sample of 10,000 individuals randomly selected from the NHS administrative register, using a variety of threshold event rates.

Results: Analysis of GP records showed that 99% of events occurred at the computed practice of registration. Sensitivity was measured as the percentage of all events that occurred within the computed periods of presence in the data. This varied based on the threshold event rate used; it was 93% for a 10% threshold and 84% for a 30% threshold. Specificity is difficult to measure; if no events are recorded, it could either indicate missing data or an individual not attending the GP. An approximation was made by counting the portion of person-months with any events recorded. For known good periods, 85% had events. Using a 10% threshold, this dropped to 75%. The ideal threshold can be selected on a per-study basis.

Conclusions: A standardized method for solving this common problem will allow for faster development of studies using this data set. Using a rigorous, tested method of verifying presence in the study population will also increase the quality of research. These methods could be beneficially applied to any linked data set that is used for more than one research project.

Corresponding author email: d.s.thayer@swansea.ac.uk

Applying missing data methods to routine data: A prospective, population-based register of people with diabetes.

Read, S H, Centre for Population Health Sciences, University of Edinburgh
Wild, S, Scottish Diabetes Research Network Epidemiology Group
Lewis, S, Edinburgh MRC Clinical Trials Methodology Hub

Background:
Routinely collected data could be used to make RCTs more efficient, either for collection of outcome data or to enhance recruitment. Unfortunately, the use of routine data in RCTs has been limited by concerns surrounding data quality and in particular, missing data. Routinely assembled data are particularly susceptible to missingness and the Scottish Care Information - Diabetes Collaboration (SCI-DC) register, a dynamic population-based register of people with a diagnosis of diabetes in Scotland is no exception.

Numerous approaches for handling missing data are available, each of which make important assumptions regarding the mechanism by which the missing data occurred. Using a 2008 extract of the SCI-DC register, we compared the use of four methods for handling missing patient BMI data in a retrospective cohort study of the association between body mass index (BMI) at date of diagnosis of diabetes and all-cause mortality in patients with Type 2 diabetes.

Methods:
The appropriateness of selected missing data approaches was investigated by assessment of the likely missing data mechanism. Descriptive analyses and logistic regression were used to identify differences in characteristics between people with (n=99,472) and without available BMI (n=117,048). Complete case analysis (CCA), population mean imputation (PMI), stochastic imputation (SI) and multiple imputation (MI) methods were applied to deal with missing data in the BMI variable. Cox proportional hazard model coefficients for the association between BMI and all-cause mortality were compared for each missing data method.

Results:
There were 41,664 deaths among the diabetes cohort between 2001 and 2008. Patients with a missing BMI were considerably more likely to have an earlier year of diagnosis (Odds ratio “Before 1995” vs. “After 2004”: 60.29 [95% confidence interval: 57.23, 63.51]). In addition, based on smoking prevalence, Charlson comorbidity indices and HDL-cholesterol data, patients with missing data also appeared to be healthier than patients without missing data.

CCA produced a J-shaped relationship between patient BMI and all-cause mortality. However, findings from PMI indicated that CCA underestimated the survival in this population. Estimates obtained from SI and MI flattened the observed J-shaped curve, though imputations were based on poor predictions.

Summary:
Despite CCA yielding results which correspond well with earlier general population studies, it is likely that our estimates were biased. In the presence of missing data where the mechanism of missingness is unlikely to be MCAR, two appropriate methods for handling missing data should be applied and results from these methods should be compared.

Corresponding author email: S.h.read@sms.ed.ac.uk

Extracting information for research from free text in electronic health records

Shah, AD, Clinical Epidemiology Group, Department of Epidemiology and Public Health, UCL, London, UK

Electronic health record systems in the UK and elsewhere contain much information as unstructured free text rather than in coded form. Free text may contain clinical information which can better characterise participants in a research study or identify outcomes which are not coded. However, free text is not easy to use in research and access may be limited because of concerns over confidentiality. For example, the UK Clinical Practice Research Datalink (CPRD) requires free text to be anonymised by CPRD staff (at a cost) before it can be released to external researchers.

Analysis of free text can be a time-consuming process if the text has to be reviewed manually. There has been much interest in the development of computer algorithms to analyse free text and convert it into a structured form which is suitable for research. Such techniques may also be useful in assisting data entry and improving the quality of information in electronic health records.

This session will discuss approaches to the automated analysis of free text.

Corresponding author email: a.shah@ucl.ac.uk

Methods for improving the estimation of multilevel survival models for large linked datasets

Stewart, CH, MRC/CSO Social and Public Health Sciences Unit, Glasgow, UK
Nobile, A, University of Glasgow, School of Mathematics & Statistics, Glasgow, UK
Titterington, DM, University of Glasgow, School of Mathematics & Statistics, Glasgow, UK
Leyland, AH, MRC/CSO Social and Public Health Sciences Unit, Glasgow, UK

Background

Linked datasets provide the ability to follow individuals over long periods of time and to study rare outcomes. Multilevel survival models should be used to account for the correlation of survival times for individuals living within the same geographical location when the outcome is time until some pre-defined event. Although multilevel survival models have been developed, computational requirements mean their use, within popular statistical software packages, is limited for large datasets. This is problematic since linked datasets are typically large, but as linkage of survey data with hospitalisation and death records enables potential risk factors for the outcome to be incorporated into models, it is important to overcome these limitations. Additional computational problems can occur if a high proportion of observations are censored. We consider the use of re-parameterisation and parameter expansion for improving the Bayesian estimation of a multilevel survival model in the presence of extreme censoring.

Data & Methods

Using 1995 and 1998 Scottish Health Survey (SHeS) data linked with death and psychiatric hospital records, we analysed time until first psychiatric admission. The linked dataset contained 15,305 individuals, aged 16-74, nested within 624 postcode-sectors in Scotland. Individuals were followed up from survey interview until 2004. Multilevel Weibull proportional hazards models were fitted in WinBUGS. Problems of correlation and slow convergence were addressed by comparing an original Weibull model, a re-parameterised Weibull model and a re-parameterised Weibull model incorporating parameter expansion. Effective sample sizes (ESS) and computing times were used to compare models after 300,000 iterations following a burn-in of 200,000 iterations.

Results

Only 137 individuals (0.9%) experienced a first-ever psychiatric admission during follow-up, implying a high proportion (99%) of censored observations. The ESSs for the original Weibull model were poor (140, 22 and 306 for the intercept, shape and higher-level variance parameters respectively). For the re-parameterised model the ESSs increased to 524, 191 and 381 for the intercept, shape and higher-level variance parameters respectively. Computing times for the original (~34.5 hours) and re-parameterised models (~38.5 hours) were similar, but further iterations were still required to achieve convergence. Incorporating the parameter expansion only improved the ESSs to 546, 204 and 434 for the intercept, shape and higher-level variance parameters respectively, but almost doubled computing time. Simulation studies revealed that such techniques worked best when fewer than 90% of the observations were censored.

Conclusions

Provided that computing time is not important, re-parameterisation and parameter expansion techniques can be effective in improving, even in the presence of extreme censoring, Bayesian estimation of multilevel Weibull survival models fitted to large linked datasets.

Corresponding author email: cstewart@sphsu.mrc.ac.uk

Consent to Data Linkage in the Context of the Avon Longitudinal Study of Parents And Children (ALSPAC): A Qualitative Study

Kennedy, M, University of Bristol
ter Meulen, R, University of Bristol
Heeney, C, Conseco Superior de Investigacion (CSIC) Madrid
Macleod, J, University of Bristol

Linkage to administrative data is becoming increasingly more wide-spread in health and social research. Facilitated by technological advances and encouraged by funders, linkage is arguably a cost-effective and efficient way to enrich research with data that is routinely collected and already exists whilst posing low risks to individuals. As with all population level research, obtaining informed consent from vast numbers of individuals can be difficult and costly. Investigations into patterns of consent to data linkage have found that although the majority of individuals are happy to consent to linkage, individuals from certain social groups are potentially more likely to refuse. Problems with consent have led to calls for it to be waived for data linkage research on the basis that it is low risk and because the potential for consent bias is damaging to research. However informed consent is enshrined in research ethics and pivotal to research ethics guidelines, protecting the interests of research participants by respecting their right to autonomy.

This paper will report some of the findings from a qualitative study in the context of the Avon Longitudinal Study of Parents And Children (ALSPAC), an on-going prospective longitudinal birth cohort study based at the University of Bristol in the UK. ALSPAC is asking young people who are members of their cohort to allow the study to link to some of their administrative data: medical, criminal, education and benefits and earnings records. Qualitative interviews have been conducted with researchers, members of the ALSPAC Ethics and Law Committee (AE&LC) and young people who have either consented to all of the data linkage requests or who have refused linkage to one or more of the records. In talking to some of the key actors in this context, I have sought to gain a contextualised understanding of the setting, explore different perspectives with regard to linkage and identify ethical issues pertinent to this context.

Consent was an important issue to all research participant groups interviewed. The members of the ALSPAC cohort have expectations regarding the consent processes of the study and the findings emphasise the importance of communication and consent in developing trusting relationships, subsequently influencing people’s willingness to participate in data linkage research in this context. However both the research ethics committee members and researchers were concerned as to how “informed” consent in the ALSPAC context can realistically be due to a number of factors including the complexity and broadness of the consent required for this type of research and the special relationship between the study and many of their participants, potentially meaning that these members of the cohort are more likely to agree to study requests. This raises an interesting ethical issue concerning the emphasis upon informed consent in research ethics guidelines; however despite the fact that consent may not be fully “informed”, it is valued by participants. Also ethically significant is that trust appears to be very influential in shaping participation with some of the members of the cohort; therefore we should not neglect the importance of informed consent and maintaining responsibility to inform participants adequately and obtain their consent responsibly.

Corresponding author email: mari-rose.kennedy@bristol.ac.uk

Patterns in consent to study enrolment and linkage to health and administrative records; evidence from the PEARL consent campaign

Davies, AJV, University of Bristol
Boyd, AW, University of Bristol
Cornish, R, University of Bristol
Macleod, JM, University of Bristol

During childhood the Avon Longitudinal Study of Parents and Children (ALSPAC), a large birth cohort study, sought consent for index children to participate using the consent of their parent/guardian. Upon reaching the age of maturity (age 18) ALSPAC sought their renewed consent to participate in the study. This consent was sought as part of the Project to Enhance ALSPAC through Record Linkage (PEARL) consent campaign. PEARL also sought consent for follow-up through collecting data from health and administrative records; including records of the participant's health, education, benefit and incomes and any criminal convictions and cautions they may have. In this paper we describe the patterns in consent outcomes and evidence as to whether the process had a differential response by participant socio-economic position and/or study participation history.

ALSPAC further sought exemption from consent requirements to link to the health records of non-responders from the National Information Governance Board, the body advising the Secretary of State for Health on these issues. Provisional exemption was granted in 2012, but only for those individuals explicitly advised of the implications of their non-response in this regard. This advice wasn't included in the original consent information sent during 2011. The influence of this change was assessed. This included any influence on response and assessed the impact it may have had on response and refusal rates from that of the revised consent materials sent in 2012.

Corresponding author email: a.davies@bristol.ac.uk

Is the use of linked routinely acquired NHS data for pharmacovigilance in children acceptable to parents and young people?

Scobie-Scott, E, Centre for Academic Primary Care and Department of Child Health, University of Aberdeen
Bond, CM, Centre for Academic Primary Care, University of Aberdeen
Shaw, DM, Institute for Biomedical Ethics, University of Basel
Helms, PJ, Child Health, University of Aberdeen

Introduction

Under reporting of adverse drug reactions (ADRs) using the UK Yellow Card Scheme, particularly in children, may delay their identification and impact on the quality of prescribing. Linking routinely collected anonymised NHS data could provide an alternative complementary system of pharmacovigilance.

Aims

The aim of the work reported here was to describe the knowledge and attitudes of young people and adults to the linkage of NHS data for pharmacovigilance. This study is part of the Child Medical Records for Safer Medicines (CHIMES) research programme.

Methods

Adults living in Scotland caring for at least one child (<16 years) (parents/guardians) and young people (14-16yrs) were selected by Research Now, a commercial research company, from their database. There was a target recruitment of 100 adults and 100 young people. A web-link was emailed, inviting participants to complete an online questionnaire which collected information on demographics, opinions of ADR monitoring, general use of online services, and data linkage for pharmacovigilance.

Results

A total of 947 adults and 1441 young people were approached. 145 and 132 respective valid responses were received (response rate 15.3% adults and 9.1% young people). More adults (44.8%) than young-people (21.2%) were aware linked NHS data is used to monitor the population's health. The majority agreed that it would be acceptable to link data for pharmacovigilance (84.9% adults and 73.5% young-people). Both groups indicated that the use of partial post codes, DOB, age and sex are acceptable fields to link but not addresses. The majority of adults (75.2%) reported it was important for people to give consent to the linkage and disagreed (61.4%) with the statement, "There is no need to ask for consent to use combined (linked) anonymised health data". Similarly, 70.4% of young people disagreed with this statement. There was less consensus on the way consent was obtained, with 24% of both groups selecting individual consent, but 28.3% of adults and only 10.6% of young people favouring an opt out process. Just under a third of adults (31.7%) thought both parents and children should be involved.

There was most support for the NHS to be responsible for the combined data (83.9% adults and 87.1% young-people) compared to an independent guardian or a private company.

Conclusion

Although the majority support the use of linked data for pharmacovigilance the preference for opt in than opt out poses challenges for the use of linked health data. Given that the CHI, which includes full addresses in the dataset, would be the unique identifier underpinning data linkage, concerns about use of this information need to be addressed.

Corresponding author email: emmascobie@abdn.ac.uk

Involving consumers in the work of a data linkage research unit

Jones, KH, Swansea University
McNerney, CL, Swansea University
Ford, DV, Swansea University

Introduction: The Health Information Research Unit conducts health-related data linkage research using anonymously-linked, routinely collected data in the Secure Anonymous Information Linkage databank. We have established a Consumer Panel to provide a patient and public voice in this rapidly growing area of work. We describe and reflect on the work of the Panel during the first year, and the feedback from the Panel members, and show how they are inputting into plans for the future.

Methods: The Panel meets quarterly and following the first meeting, initial feedback was obtained from a sample of Panel members. A review of Panel activities was carried out after a year and all members were invited to provide their views via a questionnaire survey using structured and free-text responses.

Results: The Panel was formalised by terms of reference and the election of a Chairperson from among the consumers. There are currently 10 consumer members (4 men and 6 women) from across Wales, with a range of health-related areas of interest. The Panel has been involved in a wide variety of activities in its first year. The initial feedback was tentatively positive, and the questionnaire survey identified practical measures for improvement and future work.

Conclusions: We have found the Consumer Panel to be a valuable addition to our work in the rapidly growing area of data linkage research. The views of Panel members provide a positive outlook and a fresh, and sometimes unexpected, perspective on various issues. The survey gave us useful feedback to focus our work in the forthcoming year. The lessons we have learned, and our experience of involving the Panel in various aspects of our work, may be of value to others seeking to work with consumers in data linkage research, to researchers in general, and to consumers themselves.

Corresponding author email: k.h.jones@swansea.ac.uk

Linkage of routine data to generalise results from randomised controlled trials

Harron, K, Institute of Child Health, UCL
Wade, A, Institute of Child Health, UCL
Muller-Pebody, B, Public Health England
Goldstein, H, Institute of Child Health, UCL
Gilbert, R, Institute of Child Health, UCL

Objectives

Routine data can be used to identify patients for randomised controlled trials (RCTs) but can also identify wider populations for whom trial results might be generalised. CATCH (CATheters in CHildren) is an RCT to determine the effectiveness of impregnated central venous catheters (CVCs) compared with standard CVCs for preventing blood stream infection (BSI) in paediatric intensive care (PICU). Understanding the variation in absolute risk differences in relation to the cost of purchasing CVCs for individual units could enable targeted improvement. Generalisation of results to groups not included in the trial requires information on infection rates according to case mix, taking into account on-going improvements in infection control across the NHS. Baseline infection rates are currently unknown and cannot be generalised from the trial, as participation in CATCH itself may alter reported rates. Linkage of routine data was used to generalise trial results by identifying children like those enrolled in CATCH and estimating the baseline risk of BSI for both those included and not included in the trial.

Methods

PICU admission data (PICANet) for all PICUs in England and Wales from 2003-2012 were linked with laboratory surveillance data collected by Public Health England (formerly the HPA) using match probabilities and prior-informed imputation. Poisson regression was used to estimate the baseline risk of PICU-acquired BSI (BSI occurring >48 hours following admission), adjusting for significant risk factors. As trial results are not yet published, a plausible range for the relative-risk of infection was derived from a recent meta-analysis (Gilbert,2008). The estimated absolute difference in BSI, given the use of impregnated CVCs, was derived using the estimated relative-risk range and baseline risk at the close of trial recruitment.

Results

Linked PICANet-surveillance data showed adjusted PICU-acquired BSI rates falling from 9.9 to 3.1 per 1000 bed-days between 2003-2010. Assuming this trend continued, the predicted baseline rate for the end of 2012 was 1.2 per 1000 bed-days (95% CI 0.8-1.6). Assuming that the majority of PICU-acquired BSI occurred in children with CVCs, a relative-risk of between 0.06-0.28 for BSI with impregnated versus standard CVCs would correspond to a potential absolute decrease of 0.48-0.86 BSI per 1000 bed-days using impregnated CVCs. Results will be updated with risk-factor data to the end of recruitment.

Conclusions

Linkage of routinely collected data can help determine the generalisability of trial results, supports the translation of important research findings to practice and can be used to monitor implementation into units most likely to benefit.

Corresponding author email: katie.harron.10@ucl.ac.uk

Using routinely collected data to enhance long term follow up data: an example from the Building Blocks trial.

Cannings-John , RL , South East Wales Trials Unit , Cardiff University
Robling, MR, South East Wales Trials Unit , Cardiff University

Randomised controlled trials (RCTs) are widely accepted as the gold standard to assess outcomes and safety of medical or complex interventions. Non-responders or participants lost to follow-up can be problematic in trials especially those collecting long-term self-reported follow-up. This can result in missing data and will inherently incorporate bias in trial results. In such circumstances, linkage to routinely collected data can enhance the data held by RCTs and could help in several ways. Building Blocks is a multi-centred individually randomised controlled trial evaluating the effectiveness of the Family Nurse Partnership (FNP) programme in 18 sites in England. FNP is a home visiting intervention aiming to address social exclusion and health disadvantage. The study will evaluate the effectiveness of FNP and usual care versus usual care for nulliparous pregnant women under that age of 20, recruited by 24 weeks gestation and followed until the child's second birthday. Self-reported data (through telephone and face-face interviews) has been collected from participants at baseline, 34-36 weeks gestation, 6, 12, 18 and 24 months following birth.

To enhance this self-/maternal reported data, routine data will be collected from a number of different sources: maternity, abortions, primary and secondary care. For certain outcomes such as hospital episodes of injuries and ingestions, it will be the primary source of data. The advantage of routine data in the Building Blocks trial is that this population of teenage mothers is a mobile and potentially vulnerable population and that increases the difficulty of follow-up. Therein lays the associated advantage of using routine data sources in that outcomes are verifiable and are not just based on self- / maternal report.

This presentation will examine the use of data linkage for trial data to potentially enhance data quality, and to reduce costs over face-to-face data collection. It will document the processes of linking to different sources of routine data such as the governance issues surrounding gaining adequate consent for data linkage, linkage methods and quality control regarding linkage. We will also discuss optimising linkage in the way we collect trial data (e.g. using automatic identifier checks at point of trial data collection - such as post code look-up files), the potential for establishing broader linkage to other data sets and over longer periods of time.

This is an independent report commissioned and funded by the Policy Research Programme in the Department of Health. The views expressed are not necessarily those of the Department.

Corresponding author email: canningsrl@cardiff.ac.uk

Use of electronic health records to implement a cluster randomised trial in primary care.

Dregan, A, King's College London, NIHR Biomedical Research Centre, Guy's and St Thomas' NHS Foundation Trust and King's College London
Gulliford, M, King's College London, NIHR Biomedical Research Centre, Guy's and St Thomas' NHS Foundation Trust and King's College London
Van Staa, T, Clinical Practice Research Datalink (CPRD), London, UK
McCann, G, Clinical Practice Research Datalink (CPRD), London, UK
Yardley, L, Southampton University, UK
McDermot, L, Southampton University
Ashworth , M, King's College London
Little, P, Southampton University
Moore, MV, Southampton University

Background: Electronic health records (EHRs) represent a potentially valuable resource for evaluating health interventions. This research aimed to evaluate the feasibility of implementing a cluster randomised trial in a primary care database. The trial intervention aimed to reduce antibiotic prescribing in respiratory tract infections (RTIs).

Methods: The trial was implemented using the Clinical Practice Research Datalink (CPRD). All CPRD general practices in England and Scotland were invited to participate. Allocation was stratified by region/country and practice list size. Non-trial practices provided an external comparison group. Interventions were electronic reminders, activated during consultations for respiratory infections by persons aged 18 to 59 years, installed remotely at intervention practices. Outcomes at 12 months were evaluated through analysis of EHR data collected into CPRD with linked HES. The primary outcome was the proportion of consultations for RTI with antibiotics prescribed.

Results: There were 104 (53 Intervention, 51 Control) CPRD general practices were recruited to the study. Individual participants were Intervention, 294,259; Control, 264,065; and External Comparison, 518,315. Interventions were successfully installed at all intervention practices. Banners offering accessing to the intervention material were regularly viewed, but prompts were actively viewed at only 47 per 1,000 consultations for sore throat. The median (IQR) practice-specific proportion of RTI consultations with antibiotic prescribed were: Intervention, 54 (46 to 57); Control, 54 (45 to 59); External comparison 54 (45 to 62).The linkage of CPRD with HES suggested no increase in hospital admissions following the intervention.

Implications: Cluster randomised trials may be implemented successfully, with large numbers of participants, using a primary care electronic database. Remote delivery of interventions to practices is feasible but future development of more effective interventions is required.

Corresponding author email: alexandru.dregan@kcl.ac.uk

The importance of core outcome sets for practitioners, patients, policy makers and researchers

Williamson, PR, MRC HeRC North

There is growing recognition that insufficient attention is paid to the outcomes measured in clinical trials, clinical recordkeeping in daily practice and audit. Difficulties caused by heterogeneity in outcome measurement and reporting often hinder or even prevent the comparison or synthesis of evidence in healthcare. Much would be gained if an agreed core outcome set (COS) of a minimum number of appropriate and important outcomes was measured and reported in clinical trials and routine health records in a particular condition. This does not imply that outcomes in a particular area should be restricted to those in the COS, and it is fully expected that other outcomes as well as those in the core set will continue to be investigated.

For findings to influence policy and practice, chosen outcomes need to be relevant to those making decisions about healthcare, including health service users themselves. The need to involve multiple stakeholder groups to determine what should be core means that more appropriate outcomes are likely to be measured. Examples exist where the involvement of patients in COS development has identified an outcome that might not have been considered by practitioners alone.
The COMET (Core Outcome Measures in Effectiveness Trials) Initiative brings together people interested in the development and application of COS, including health professionals, health service users, trialists, trial funders, regulators, systematic reviewers, policy makers and journal editors. COMET (www.comet-initiative.org) aims to provide guidance on developing COS, methods to include user involvement, and preparing reporting standards for such projects.

The COMET Initiative has developed a database of studies relevant to core outcome sets for use in clinical trials, where researchers, patients and the public can access information on outcomes recommended for use in clinical research and practice (www.comet-initiative.org/studies/search). In the months since the launch of the COMET website and database (August 2011), 2469 searches have been undertaken, with 7787 individuals (13524 visits) from 104 countries visiting the site.

COS will increase the efficiency and value of the research process. Design of new trials will be simplified, risk of measuring inappropriate outcomes reduced, and selective reporting of outcomes less likely. Trial conduct will be more efficient if the key outcomes for effectiveness trials are those routinely recorded in the electronic health record. By improving the evidence base, COMET will make it easier for people to make well-informed decisions about healthcare.

Corresponding author email: prw@liv.ac.uk

Using G-Cloud services to build a secure record linkage service

Cortina-Borja, M, UCL Institute of Child Health
Hutchinson, R, UCL Institute of Child Health
Thomas, A, UCL Institute of Child Health
Langan, P, AIMES Grid Services CIC Ltd
Messenger, M, AIMES Grid Services CIC Ltd
Kehoe, D, AIMES Grid Services CIC Ltd
Castillo, FD, UCL Institute of Child Health

Access to healthcare data in the UK is carefully controlled to protect the interests of patients. In normal circumstances informed consent must be obtained from patients or relatives and the research project protocol must obtain appropriate research ethics approval for researchers to access identifiable or sensitive clinical information. Section 251 (S251) of the NHS Act 2006 allows the common law duty of confidentiality to be set aside in specific circumstances where anonymised information is not sufficient and where patient consent is not practicable. Applications for S251 support are currently considered by the Ethics and Confidentiality Committee of the National Information Governance Board (NIGB).

All NHS and partner organisations in England are assessed annually using an "Information Governance Toolkit", derived from the ISO-27001 information security standard, providing a mechanism for self-directed audit of information security and governance and requires the participating organisations to provide appropriate evidence of their compliance. Universities commonly lack sufficient institutional information security management systems to meet the demands of NHS Information Governance and would benefit from the option to access secure third party information services.

Our project, funded by TSB in partnership with AIMES Grid Services (a G-Cloud service provider) project has established a secure record linkage environment. The project developed a demonstrator application for enhancing data from the unlinked anonymous (UA) newborn anti-HIV surveillance programme in North Thames at the Institute of Child Health (ICH) with the Office for National Statistics (ONS) data matching for the provision of services for the Health Protection Agency (HPA). Re-engineering of the record linkage workflow has the potential to improve the data quality assurance and allow other forms of UA surveys (e.g. on Hepatitis C) to be developed on the basis of the HIV surveillance programme.

The resulting environment has been successfully certified to the ISO-27001 information security standard and complies with Level 2 of the NHS Information Governance Toolkit.

Corresponding author email: t.castillo@ucl.ac.uk

Overcoming complexity and tedium: an object-oriented programming framework for linked data researchers.

Churches, T, The Sax Institute, Sydney, Australia
Fox, D, The Sax Institute, Sydney, Australia

Many researchers analysing linked, routinely-collected health data report spending a great deal of preparatory time on basic data quality assurance and manipulation of the source data into a form suitable for statistical analysis. The complexity and difficulty of these essential tasks tend to increase multiplicatively as the number of different linked data sources used in a study rises, and as the complexity of study designs grow. In addition, many researchers use statistical packages such as SAS or SPSS to carry out this data manipulation prior to analysis. The programming paradigms which underlie these statistical packages date from the 1970s, and thus many researchers do not have access to more modern ways of writing computer programs.

To address these issues, we have created an object-oriented programming framework specifically designed for use by population health and health services researchers working with large, linked health data files. The framework uses free, open-source software components (the Python programming language, the PostgreSQL database, and the R statistical analysis environment) to allow researchers to quickly set up and execute very complex study designs using linked, already-collected data. The framework abstracts the underlying details of data files, file merges, and other low-level programming tasks, and instead allows the researcher to focus on, and program in terms of, epidemiological concepts such as case and referent sets, or person-age cohorts, or just sets of chronologically ordered health events for each candidate person in a study. The result is a much terser but more readily understood body of program code for each study, potentially an order of magnitude smaller than equivalent traditional data preparation code.

The design and implementation of the framework will be briefly described, and a complete set of data preparation steps for a moderately complex survival analysis study will be demonstrated. The framework is readily extensible in a generic fashion by the researchers using it, thus facilitating code re-use between studies. It also facilitates the use of formal software testing methods, allowing researchers to be much more confident that the program code which they have written is doing exactly what they intended.

At this stage, the framework primarily addresses time-consuming and often challenging data preparation steps and other pre-analysis data manipulation and checking tasks. The resulting prepared data files can then be used with whichever statistical package the researcher prefers. However, prospects for extension of the framework to cover routine and stereotypic statistical analysis and model fitting tasks will also be canvassed. We believe that programming frameworks such as this have the potential to dramatically improve researcher productivity, allowing linked data studies to be undertaken much more quickly and cheaply, allowing researchers to focus on epidemiological and statistical issues rather than getting bogged down in the complex but tedious programming tasks needed to prepare their data for analysis.

Corresponding author email: tim.churches@saxinstitute.org.au

An Introduction the SAIL Databank

Jones, C, Swansea University
Ford, DV, Swansea University
Jones, KH, Swansea University
D'Silva, R, Swansea University
Thompson, S, Swansea University
Brooks, C, Swansea University
Heaven, ML, Swansea University
McNerney, CL, Swansea University
Lyons, RA, Swansea University

The SAIL Databank was established in 2006 funded by NISCHR to realise the potential of anonymised data for health research.

The objectives for the SAIL Databank are: 1) ensuring data transportation is secure; 2) operating a reliable record matching technique to enable accurate record linkage across datasets; 3) anonymising and encrypting the data to minimise the risk of individual re-identification; 4) applying measures to address disclosure risk in data views created for researchers; 5) ensuring data access is controlled and authorised; 6) scrutinising proposals for data utilisation and approving output; and 7) gaining external verification of compliance with Information Governance. Having successfully met these objectives the SAIL Databank is a research ready platform for the population in Wales supporting research projects with a value of over £35 million.

This presentation will introduce you to the SAIL Databank describing the methods employed to securely populate a central anonymised and linkable data repository, the governance around applying to use the SAIL Databank, and the technology used to provide secure remote access to the SAIL Databank.

Corresponding author email: c.jones@swansea.ac.uk

Role of the Research Co-ordinator

Campbell, F, ISD Scotland
Nogueira, R, ISD Scotland

The ScottisH Informatics Programme (SHIP) infrastructure provided by ISD supports research by providing expertise in study design, facilitating research approvals and providing datasets, including linked data, in a secure environment. Our work supports researchers to analyse and interpret complex datasets, answering key questions about our society. To deliver SHIP, ISD operates a state-of-the-art technical infrastructure managed by the ATOS Origin Alliance. This environment provides a high powered computing service, secure analytic environment, secure file transfer protocol, and provision of a range of analytic software (SPSS, STATA, SAS and R).

Research coordinators:

- Help researchers to finalise their study design by providing expert advice on study feasibility. Using our extensive knowledge of health datasets we assist in defining what data are available coupled with relative strengths and weakness of datasets and their fitness for the proposed study
- Facilitate completion of the permissions (e.g. ethics, Privacy Advisory Committee, CHI Advisory Group, Caldicott Guardians Forum)
- Liaise with all parties involved including data suppliers (both external to ISD and within ISD), the ATOS safe haven and ISD's Indexing Service to provision data within the safe haven

If required, provide analytical capacity to undertake the analyses on behalf of customers. ISD have been piloting the SHIP infrastructure and the research co-ordinator role since early 2013. This talk will review the first six months of the research co-ordinator role and explain how the co-ordinators will support researchers in practice.

Corresponding author email: fiona.campbell2@nhs.net

Exploring Social Segregation through urban form indicators and existing health data

Pasino, p, University of Strathclyde

The link between social segregation and ur.ban form has widely been debated in the existing literature, with scholars such as Laura Vaughan and Lars Marcus for example (Vaughan 2005; Vaughan, Clark et al. 2005; Marcus 2007; Vaughan 2007) studying Robert Booth’s maps of Poverty in 19th century London through the application of Space Syntax, a set of theories and techniques able to objectively assess the physical and spatial attributes of urban settlements in relation to patterns of human activity, to argue that segregation is inherently a spatial problem.

This paper further investigates the subject through the combination of an objective method of analysis of the structure of the city, the Multiple Centrality Assessment (Porta, Crucitti et al. 2005), which considers Centrality as a critical element of the structure of all complex networks, and the study of the distribution of drug abuse in the form of Methadone Prescription occurrence as a proxy indicator of deprivation (Walsh, Bendel et al. 2010).

The analysis links existing health data (collected by the NHS) to urban morphology data developed by the author through the MCA ,shifting the field of research from the historical to the contemporary city; the paper will also discuss the ethical and data handling challenges, which were resolved through Scottish Informatics Programme (SHIP).

Findings will shed new light on the hypothesis of a link between social deprivation and physical exclusion, and will be illustrated through statistical analysis interpolating urban morphology indicators and the occurrence of methadone prescriptions.

Corresponding author email: pasino.paola@strath.ac.uk

Exploring the multidimensional influence of access to care on potentially preventable hospitalisations

Falster, MO, Centre for Health Research, School of Medicine, University of Western Sydney
Leyland, AH, MRC/CSO Social and Public Health Sciences Unit, Glasgow
Elliott, RF, Health Economics Research Unit, University of Aberdeen
Jorm, LR, Centre for Health Research, School of Medicine, University of Western Sydney

Background

Potentially preventable hospitalisations (PPH) are those which could potentially be prevented through timely access to quality primary and preventive care, and are used as an indicator of health system performance. Various basic measures of "access", including self-reported access to care, rates of physician supply, and access to community health centres, have been shown to be associated with PPH. However "access" is a complex negotiation of individual health and socio-economic barriers, localised allocation of resources, and structural configuration of health services. This study sought to operationalise a model for jointly examining the influence of individual and structural forms of "access" on PPH, using linked routinely collected data and multilevel modelling.

Methods

The study included 267,961 participants in the 45 and Up Study from New South Wales (NSW), Australia. Baseline personal characteristics were obtained from a self reported questionnaire. Linked hospital morbidity data were used to identify PPH admissions in follow-up, and linked claims data from Medicare, Australia's universal health insurance provider, were used to identify use of primary care services. Measures of "access" were identified from linked and unlinked data. Personal measures of "access" included concessional health insurance status, health-care seeking behaviour and having a regular provider of care. Structural measures of "access" included accessibility of the geographic area, density of the local health workforce, perceived access to primary healthcare, and hospital bed occupancy rate. Multilevel modelling was used to account for the complex structure of the data, with individuals nested within geographic areas of residence. Membership of hospital catchments was constructed using patterns of linked hospital admissions.

Results

Among the 267,961 participants, 5.9% (n=15,740) had a PPH admission. Participants were nested within n=199 Statistical Local Areas of residence, and were admitted to 326 hospitals in NSW. 20% of participants reported having subsidised care through a concession card, 53% had private health insurance, and 68% had a regular provider of GP care. The local workforce ranged from 0-165 medical workers per 10,000 population, and the average hospital bed occupancy rate varied from 31-100%. Both personal and structural measures of access made independent contributions to rates of PPH.

Conclusions

The availability of linked cohort, hospital and Medicare data, and the capacity of multilevel modelling to incorporate information at the individual, area, and service-level have allowed us to operationalise a framework for examining the complex structure of, and interactions between, components of "access" to care and PPH admissions.

Corresponding author email: michael.falster@uws.edu.au

Linking spatial accessibility of GP services to diabetes treatment

Fry, R, Swansea University
Atkinson, M D, Swansea University
Rodgers, S E, Swansea University
Grinnell, D, Swansea University

The geographic distribution of health care provision has long been associated with the Health Inequalities in Wales. According to the Association of Public Health Observatories (APHO) an estimated 9% of the Welsh population have diagnosed or undiagnosed type I or type II diabetes (Public Health Wales Observatory, 2011).

Primary care services, especially GP's, are playing an increasing role in the management and diagnosis of diabetes. Early diagnosis and management of diabetes is essential to minimising the risk of complications resulting from diabetes and therefore good access to primary care is an essential aspect of treatment.

The Audit Commission in Wales (Audit Commision in Wales, 2003) recognised that there is variation in the range of care available through general practice services at a local health board level. However, local health boards geographically, cover large areas and as such spatial access to GP services varies substantially within these administrative boundaries. Using GIS and the SAIL databank we have developed a methodology where we can anonymously link residential level accessibility scores to an individual's health records.

The pilot study area, Abertawe Bro Morgannwg University Local Health Board (ABMU), comprises of three unitary authorities Swansea, Neath Port Talbot and Bridgend. In the SAIL databank we have data for 77 GP practices in ABMU in which we have 47000 patients with diagnostic or prescription evidence of diabetes. We have HbA1c values for 83% of these. We can use HbA1c as a measure of control of diabetes. We can also assess the presence and severity of diabetic complications from the GP records and also, by linkage, from hospital inpatient records.

This paper assesses the potential health impacts (in terms of diabetes) of variability in spatial accessibility to primary care services for a small study area in South Wales. Further to this the paper examines the impact of using different accessibility measures and spatial units when developing accessibility profiles for an area and demonstrates the potential of linking spatially derived indices to routinely collected health data.

References

Audit Commision in Wales, 2003. Diabetes services in Wales, Cardiff: Audit Commision in Wales.
Public Health Wales Observatory, 2011. An estimated nine per cent of Welsh residents have diabetes. [Online] Available at: http://www.wales.nhs.uk/sitesplus/922/news/18001
[Accessed 24 4 2013].

Corresponding author email: m.atkinson@swansea.ac.uk

ONLINE INTERACTIVE ATLAS ON CHRONIC DISEASES AND MENTAL DISORDERS

Vanasse, A, Université de Sherbrooke; PRIMUS group, CRCELB (CHUS)
Courteau, J, PRIMUS group, CRCELB (CHUS)
Courteau, M, PRIMUS group, CRCELB (CHUS)

Context: In order to transfer complex epidemiological and clinical results into available, relevant and useful information for decision-makers, there is a need to merge disparate data (medical administrative, populational, demographic, geographic and spatial data) from different sources.

Objective: To develop an online interactive atlas on chronic diseases and mental disorders.

Methods: The Atlas acquires and combined medical administrative, populational and spatial data from the health ministry and national statistical agencies. Exhaustive cohorts of patients diagnosed with chronic diseases (cardiovascular risk factors, chronic pain) and mental disorders (schizophrenia, mood disorders) in Quebec (Canada) between 2000 and 2007 was built using the provincial medico-administrative database. Information on the prevalence, mortality, morbidity, utilization of medical resources and treatment was aggregated at different socio-geographical levels, such as local health regions, rural/urban populations, material or social deprivation neighbourhoods. The atlas used an Extract Transform and Load (ELT) process to populate a datacube with six dimensions. The Atlas was built using the JMap/Solap technology (K2 Geospatial/Intelli3).

Results: The Atlas on chronic diseases and mental disorders is an online, easy-to-use health information system that allows users, particularly those concerned by health monitoring, resource allocation and planning, to interact with aggregated health information at different spatial and population units of analysis and to produce tables, graphs or maps almost instantly. As an example, users can observe a strong variation - by a factor 6 - in the prevalence of schizophrenia in 2004-2005 between neibourhoods classified according to a social and material deprivation index; the Atlas also shows high rates of emergency utilization in rural areas, both for men and women with chronic pain and a net gradient of over-utilization for both ambulatory and emergency cares with deprivation, especially social deprivation. Cross-analysis by age group and sex also reveal high rates of over-utilization in patients over 65 with chronic pain.

Conclusion: By providing health related knowledge on regions or specific subpopulations (gender, rural/urban, deprivation level), the Atlas is an excellent tool to detect spatial inequalities that would have been overlooked using traditional information system. This kind of Atlas may influence the priority-setting and resource allocation based on empirical knowledge.

Corresponding author email: alain.vanasse@usherbrooke.ca

Estimation of familial effects on hospitalisation for common childhood infections

de Klerk, N, Telethon Institute for Child Health Research
Burgner, D, Murdoch Children's Research Institute
Colvin, L, Telethon Institute for Child Health Research

One of the most powerful uses of data linkage is in combining detailed data from one study to routine data from administrative datasets to form new cohort studies. We have carried out two separate studies using both twin and sibling population linked datasets to investigate hospital admissions for infectious disease in siblings and in twins, and ENT surgical procedures in twins. The sibling analyses included hospital admissions from 1980-2000 and the twin analyses included pairs born 1980-1992, with all hospital admissions until 2000.

(1) Sibling analyses of hospitalisation with infection:
We used observed and expected admission rates based on groupings of age and sex. We found strong unadjusted sibling concordances for common infections, with a sibling risk ratio (λ s) for admission with the same infection type of 1.9 (95% CI 1-9-1.9), but no increased risk in siblings for admission with a different infection types (λ s = 0.9, 95% CI 0.9-1.0). This indicates that genetic determinants of hospitalisation with infection are likely to be broadly infection-specific, and hence pathogen-specific, as predicted from an evolutionary perspective. The λ s was significantly but variably increased for specific infection groups; e.g. influenza (λ s = 6.4, 95% CI 4.5-9.1), RSV bronchiolitis (λ s = 4.0, 95% CI 1.9-8.4), parasitic infections (λ s = 6.8, 95% CI 4.4-10.5) and pneumonia (λ s = 1.2, 95% CI 1.1-1.3). The variation in λ s may partly reflect relative contribution of genetic determinants to severity of infection types, with more marked genetic effects in those infections with a narrow range of causative pathogens. The median time between sibling admissions was about 3 years, implying that the increased risk in siblings was unlikely to reflect sharing of particularly virulent pathogens.

(2) Twin analyses of infection-related hospitalisations:
Using an established logistic regression approach we found strong genetic effects for: (i) ENT procedures (e.g. odds ratios (OR) of 13.4, 95% CI 6.3-29 for any ENT procedure, and OR = 4.6, 95% CI 1.8-12.2 for MVTI, with tetrachoric correlation coefficients of 0.90 for MZ vs. 0.67 for DZ) and; (ii) for specific infections (e.g. OR = 2.3, 95% CI 0.8-6.3 for bronchiolitis, and OR = 3.0, 95% CI 1.2-7.4 for systemic viral infections). These effects could have been due in part to a tendency for concurrent admissions, or different common environmental effects between MZ and DZ twins. We plan to repeat these studies with double the number of twins and longer follow-up.

Corresponding author email: nickdk@ichr.uwa.edu.au

Primary care data linkage to investigate risk factors associated with emergency hospital admission for Chronic Obstructive Pulmonary Disease

Hunter, LC, NHS Lothian
Weir, CJ, University of Edinburgh
Butcher, I, University of Edinburgh
Wild, S, University of Edinburgh
Fischbacher, CM, Information Services Division Scotland
McAllister, D, University of Edinburgh
Hewitt, N, NHS Lothian
Hardie, RM, NHS Lothian

Chronic obstructive pulmonary disease (COPD) is the third most common reason for hospital admission in Scotland and is considered to be a potentially preventable admission. The aim of this study was to investigate the association of patient characteristics and primary care interventions with risk of emergency hospital admission for acute exacerbation of COPD (AECOPD).

Methods

Primary care data were extracted from 72 (70%) of 103 eligible general practices in Lothian, Scotland for patients with a COPD diagnosis between 2000 and 2008, with follow up until 31st March 2010. The primary care data were linked to hospital admissions, spirometry, and mortality data. A time dependent Cox proportional hazards regression model was used to investigate time to first hospital admission for AECOPD. The model adjusted for age at diagnosis, sex, socio-economic status (SES), disease severity, smoking status, body mass index (BMI), previous admission for COPD, previous intervention for respiratory disease, co-morbidities, palliative care, prescriptions of statins, beta blockers and inhaled respiratory drugs. Interventions investigated included COPD indicators from the Quality and Outcomes Framework (QOF).

Results

There were 7002 people with COPD in the cohort of whom 26% had one or more hospital admissions during 4.4 years mean follow-up time. In the adjusted model the following baseline characteristics were significantly associated with a shorter time to first admission for AECOPD: being older, low BMI, more severe COPD, comorbidities, admission for COPD prior to diagnosis, intervention for respiratory disease prior to diagnosis. Higher SES, being an ex-smoker or never smoking and having a high BMI were associated with a longer time to admission in the fully adjusted model. None of the COPD interventions investigated were associated with a significantly reduced risk of admission in the fully adjusted model. Conversely, influenza vaccination (HR 1.20; 95%CI=1.08-1.32), inhaler check (HR 1.18; 95%CI=1.06-1.31), and pulmonary rehabilitation (HR 1.61; 95%CI=1.02-2.54) were associated with shorter time to admission, suggesting confounding by indication.

Conclusions

This study has presented several challenges, from gaining ethical approval to the need to use time dependent analysis to avoid immortal time bias. Although there are a wealth of data within the linked database, it was not possible to appropriately control for confounding in a time dependent way. This study highlights some of the potential problems in using secondary data to investigate the effect of interventions in a retrospective observational study.

Corresponding author email: leonie.c.hunter@nhslothian.scot.nhs.uk

Making sense of MS using linked data

Jones, KH, Swansea University
Ford, DV, Swansea University
Middleton, RM, Swansea University
Noble, JG, Swansea University

Introduction

The UK MS Register aims to address the need for an increased knowledge-base about Multiple Sclerosis by bringing together datasets from multiple sources to create a rich resource for research, policy development and service planning.. The aims of this paper are to: briefly describe the Register model and some of the challenges that have been encountered in its development, to present research findings and future plans.

Methods

The data to be linked are collected from information systems operating in NHS neurology clinics, from sources of routine administrative data, and directly from people with MS via a purpose-built web portal. The portal acts as a questionnaire delivery platform and includes validated patient reported outcome measures. The responses are collated with basic demographic and descriptive MS data and analysed in SPSS (v.20).

Results

Data from the neurology clinics accrue as patients attend and provide informed consent. However, data quality is variable and assurance processes are underway. The availability of routine data varies across the UK, but the Register benefits from linkage with SAIL data in Wales. Over 10,000 people with MS have registered to provide their data via the portal, yielding a large cohort for research studies, including on mental well-being, the impact of MS and quality of life. Results of this work will be presented.

Conclusion

There are many challenges in bringing together disparate types of data from multiple sources. We have been overwhelmed by the willingness of people with MS to provide their data to the Register via the portal. We have a programme of research and plans to continue developing and strengthening the Register to benefit people with MS and to support further research and service planning.

Corresponding author email: k.h.jones@swansea.ac.uk

An investigation into the use of aspirin and newer antiplatelets medications in Scotland following acute myocardial infarction

McTaggart, SA, Information Services Division, NHS National Services Scotland
Wyper, G, Information Services Division, NHS National Services Scotland
Bishop, I, Information Services Division, NHS National Services Scotland
Hurst, A, College of Pharmacy, University of Kentucky
Bennie, M, University of Strathclyde, Glasgow

OBJECTIVE: To investigate the use of aspirin and newer antiplatelets medications (NAMs: clopidogrel, prasugrel and ticagrelor) in patients following acute myocardial infarction (AMI).

METHODS: Patients having a hospital admission for AMI during 2010 were identified from Scottish morbidity reporting (SMR) 01 returns. Patients were excluded if the event was recorded as a subsequent AMI or they had any record of AMI in the preceding five years in order to minimize the effect of treatment carry-over.
Data was extracted from the national primary care prescribing information system for any prescriptions for aspirin or a NAM in the identified cohort covering the period from three months prior to the AMI until December 2012. Patients were categorized as treated with: aspirin alone; a NAM alone; both aspirin and a NAM, or: neither aspirin nor a NAM at 3month intervals for the study period. Cumulative deaths at the same time points were recorded and patients were stratified by age and gender.

RESULTS: We identified 10,602 patients discharged from Scottish hospitals during 2010 following an acute myocardial infarction and excluded 824 patients to minimise the effects of treatment carry-over leaving 9,778 patients for subsequent analysis. Mortality in patients under 70 years was 4.0% in the first three months following AMI rising to 11.3% after 18+ months of follow up. Females aged under 70 years showed higher mortality following AMI than males of the same age (RR=2.06, 95%CI: 1.65-2.56) in the first three months and this excess persisted throughout the study period. Mortality in patients aged 70 years and over was higher at 15.5% in the first 3 months rising to 49.3% by 18+ months but there was no difference between genders in this age group. In the first three months following AMI, surviving patients (n=8,801) were receiving the following medication: 58.7% aspirin + NAM; 26.7% aspirin alone; 5.8% a NAM alone; 8.8% neither aspirin nor NAM. Use of neither aspirin nor NAM was more likely in females (RR=1.56, 95%CI: 1.37-1.79) and particularly those 70 years and over (RR=2.77, 95%CI: 2.32-3.30) compared with males under 70 years. Patients over 70 years were more likely to receive aspirin alone (RR=1.45, 95%CI: 1.35-1.56) and females over 70 years were the least likely to receive the aspirin + NAM combination (RR=0.64, 95%CI: 0.60-0.67 in comparison to males under 70 years).

CONCLUSION: This study reveals differences based on both age and gender in both mortality and the use of antiplatelets medications following AMI. The reasons for these and whether they are inter-related requires further study.

Corresponding author email: stuart.mctaggart@nhs.net

Use of linked primary and secondary care data to improve incidence estimates of community-acquired pneumonia in older adults in England

Millett, ERC, London School of Hygiene and Tropical Medicine
Quint, JK, London School of Hygiene and Tropical Medicine
Smeeth, L, London School of Hygiene and Tropical Medicine
Daniel, RM, London School of Hygiene and Tropical Medicine
Thomas, SL, London School of Hygiene and Tropical Medicine

Background

Community-acquired pneumonia (CAP) is a common cause of morbidity and mortality among those aged ⩾65 years, who account for a growing percentage of the UK's population. It has been estimated that a third of pneumonia cases are hospitalised; this is likely to be considerably higher among older adults. Current incidence estimates for CAP among older adults in the UK use either primary or secondary care data, and thus do not fully capture the true burden of the disease. We compared incidence estimates of CAP using linked primary and secondary care data to those from the same patients' primary care records alone.

Methods

Electronic general practice records from the Clinical Practice Research Datalink (CPRD) were linked to Hospital Episode Statistics (HES) inpatient data, to estimate incidence of CAP among older adults in England between April 1997-March 2011, by age, sex, region and over time. CAP incidence was then estimated using CPRD data alone.

Pneumonia codes identified in CPRD or as the primary code of the first episode of a HES spell (linked data only) were regarded as part of the same illness-episode if they were within 28 days of each other. Incident cases of pneumonia were regarded as hospital-acquired if in the previous 14 days the patient had been discharged from hospital (using linked data HES records for any hospitalisation) or there was a CPRD hospital code (using the unlinked data).

Patients were considered "not at risk" of pneumonia during and for 28 days after an illness episode (both datasets), or during or in the 14 days after a HES hospitalisation (linked data only). This person-time was excluded from the appropriate analysis.

Results/Discussion

The study population comprised 917,852 patients of whom 28,976 had one or more CAP episode over the time period. CAP incidence among >65 year olds was 49% higher using linked CPRD-HES data compared to CPRD-only (9.34 vs 6.28 episodes /1000 person-years respectively). The difference in rates increased with age, over time, and varied widely by Strategic Health Authority region. Full results and exploration of trends will be presented.

In the same population, inclusion of HES data allows better identification of pneumonia diagnoses and classification of person time at risk, and this increases CAP incidence estimates substantially. In an older population (among whom hospitalisations are common), use of linked primary and secondary care data gives better estimates of disease burden.

Corresponding author email: elizabeth.millett@lshtm.ac.uk

Pathways in aged care: what we learn from record linkage

Anderson, P, Australian Institute of Health and Welfare
Dickinson, T, Australian Institute of Health and Welfare
Guiver, T, Australian Institute of Health and Welfare
Karmel, R, Australian Institute of Health and Welfare
Powierski, A, Australian Institute of Health and Welfare

Background

Coordination of aged care services is important to provide appropriate services cost-effectively. For the Pathways in Aged Care (PIAC) study a consortium centred at the Australian Institute of Health and Welfare (AIHW) linked data sets for aged care assessments, nursing home use (permanent and respite care), three major community aged care service programs and deaths. These data come from two main sources: ongoing administrative data and annual program-specific national minimum data sets. Data on health conditions leading to need for care is collected as part of the assessment program. The resulting data for the PIAC cohort allows direct investigation into people's use of aged care services over the two years after assessment.

Method

A cohort approach was taken. The study cohort was derived from the 2003-04 aged care assessment data set. The remaining data sets in the study were linked to this file using a stepwise deterministic linkage algorithm developed by the AIHW specifically for this project. The algorithm uses three measures of link quality to identify suitable linkage keys and the order in which they should be used. For PIAC, the algorithm used selected letters of name, date of birth, sex and region of residence to identify matches to the cohort.

Using the PIAC linked data, changes in program use over the two years following people's first assessment completed in 2003-04 (reference assessment) was examined. Analysis concentrated on the 33,300 cohort members who had not previously accessed an aged care program. Use of care over time and the effect of disease on program use were analysed. Concurrent use of programs was also investigated.

Results

Among 33,300 cohort members:

- There was great diversity in the programs used and the timing of program use.
- Assessments do not necessarily lead to program use: almost 25% did not use aged care services in the 2 years following their reference assessment (one-quarter of these died).
- Take-up of services was greatest in the first month after assessment. However, the timing and type of services used varied considerably with health condition and carer availability.
- After two years, of those still alive:
  - one-third were in permanent RAC;
  - almost one-quarter were using community care programs;
  - just over 40% were not accessing an aged care program in the study.

The utility of these analyses has resulted in the Institute being funded to develop a more general aged care linked database for 2003-03 to 2010-11.

Corresponding author email: tenniel.guiver@aihw.gov.au

Different effects of age, adiposity and physical activity on the risk of ankle, wrist and hip fractures in postmenopausal women: UK cohort linked to hospital admissions databases

Armstrong, MEG, University of Oxford
Cairns, BJ, University of Oxford
Banks, E, The Australian National University
Green, J, University of Oxford
Reeves, GK, University of Oxford
Beral, V, University of Oxford

Bone 2012; 50(6):1394-1400

BACKGROUND
While increasing age, decreasing body mass index (BMI), and physical inactivity are known to increase hip fracture risk, whether these factors have similar effects on other common fractures is not well established.

METHODS
We used prospectively-collected data from a large cohort to examine the role of these factors on the risk of incident ankle, wrist and hip fractures in postmenopausal women. 1,155,304 postmenopausal participants in the Million Women Study with a mean age of 56.0 (SD 4.8) years, provided information about lifestyle, anthropometric, and reproductive factors at recruitment in 1996-2001. All participants were linked to National Health Service cause-specific hospital records for day-case or overnight admissions.

RESULTS
During follow-up for an average of 8.3 years per woman, 6807 women had an incident ankle fracture, 9733 an incident wrist fracture, and 5267 an incident hip fracture. Adjusted absolute and relative risks (RRs) for incident ankle, wrist, and hip fractures were calculated using Cox regression models. Age-specific rates for wrist and hip fractures increased sharply with age, whereas rates for ankle fracture did not. Cumulative absolute risks from ages 50 to 84 years per 100 women were 2.5 (95%CI 2.2-2.8) for ankle fracture, 5.0 (95%CI 4.4-5.5) for wrist fracture, and 6.2 (95%CI 5.5-7.0) for hip fracture. Compared with lean women (BMI < 20 kg/m2), obese women (BMI ⩾ 30 kg/m2) had a three-fold increased risk of ankle fracture (RR = 3.07; 95%CI 2.53-3.74), but a substantially reduced risk of wrist fracture and especially of hip fracture (RR = 0.57; 0.51-0.64 and 0.23; 0.21-0.27, respectively).

CONCLUSIONS
Physical activity was associated with a reduced risk of hip fracture but was not associated with ankle or wrist fracture risk. Ankle, wrist and hip fractures are extremely common in postmenopausal women, but the associations with age, adiposity, and physical activity differ substantially between the three fracture sites. This study demonstrates the utility of data linkage with hospital records, for studying fracture outcomes in a large UK cohort.

Corresponding author email: miranda.armstrong@ceu.ox.ac.uk

The impact of first and second eye cataract surgery on injurious falls that require hospitalisation: a whole population study

Meuleners, L, Curtin-Monash Accident Research Centre, Curtin University, Perth, WA
Fraser, ML, Curtin-Monash Accident Research Centre, Curtin University, Perth, WA
Ng, JQ, Eye and Vision Epidemiology Research Group, Perth, WA
Morlet, N, Eye and Vision Epidemiology Research Group, Perth, WA

Background: Cataract is the leading cause of reversible vision impairment in developed countries and may increase the risk of falls in older adults due to reduced balance, stability and hazard detection.

Methods: The Western Australian Data Linkage System was used to compare the number of hospital admissions from injuries due to a fall among adults aged 60+ in Western Australia two years before first eye, between first and second eye and two years after second eye cataract surgery.

Results: Poisson regression analysis based on generalised estimating equations compared the frequency of falls two years before first eye cataract surgery, between first and second eye surgery and two years after second eye cataract surgery after accounting for the confounding effects of sex, age, marital status, and comorbidities. The risk of an injurious fall that required hospitalisation doubled (risk ratio 2.14, 95% confidence interval 1.82 to 2.51) between first and second eye cataract surgery compared to the two years before first eye surgery. There was a 34% increase in the number of injurious falls that required hospitalisation in the two years after second eye cataract surgery compared to the two years before first eye surgery (1.34, 1.16 to 1.55). There was an increase of injurious falls with increasing age with the risk in those aged 85+ almost 7 times (3.93 to 11.93) that of the 60-64 years age group. Women (2.24, 1.91 to 2.63) and those with at least one comorbidity (2.84, 2.46 to 3.29) had a significantly increased risk of falling.

Conclusions: There may be an increased risk of falls after cataract surgery which ophthalmologists should consider when discussing risks and benefits with patients. This has important implications for the timely provision of second eye surgery for older adults as well as appropriate refractive management between surgeries.

Corresponding author email: l.meuleners@curtin.edu.au

Enhancing the intensively phenotyped and genotyped Generation Scotland Scottish Family Health Study cohort through record linkage.

Linksted, PJ, University of Edinburgh
Campbell, A, University of Edinburgh
Hocking, LJ, University of Aberdeen
Smith, BH, University of Dundee
Porteous, DJ, University of Edinburgh
Padmanabhan, S, University of Glasgow
McGilchrist, M, University of Dundee

Generation Scotland: Scottish Family Health Study (GS:SFHS) is a family-based genetic epidemiology study with DNA and socio-demographic and clinical data from about 24,000 volunteers in ~7000 family groups recruited across Scotland between 2006-2011, with the capacity for longitudinal follow-up through record linkage and re-contact. Biological samples and anonymised data form a resource with broad consent for academic and commercial research on the genetics of health, disease and quantitative traits of current and projected public health importance.

All participants provided blood or saliva from which DNA was extracted, and completed a demographic, health and lifestyle questionnaire. 21,476 attended a dedicated study clinic and underwent detailed clinical assessment, including anthropometric measures, cardiovascular, respiratory, cognitive and mental health. 10,000 samples have undergone genotyping (GWAS) analysis. These measures maximize the power of the resource to explore the interplay between genes, environment and lifestyle factors in health risk (and protection) and identify, replicate or control for genetic factors associated with a wide spectrum of illnesses.

Consent for record linkage and re-contact with participants greatly enhances the resource for long-term follow-up and future studies to enable detailed analysis across a wide range of areas of interest.

With Research Tissue Bank status and its well-governed managed access process, this resource is made available to researchers both nationally and internationally, with over 100 collaborations underway or completed, including epidemiological studies in pain, cognitive function and mental health, genetic replication studies of lung function and COPD, hypothesis-driven and genome-wide association studies, through to identification of participants with genotypes or phenotypes of interest for re-contact and further study (www.generationscotland.org).

Linkage of GS study data with SMR and prescribing data held by NHS ISD is underway in several projects. These linkages will allow a wider characterisation of the cohort beyond that described in the cohort profile paper: Smith et al IJE 2012. They will allow evaluation of a number of endpoints within the linked data set both at the time of linkage and projected times in the future to support scientific planning within the resource and allow identification of confirmed endpoints for targeted future studies (e.g. as biomarker analysis for risk prediction, using stored GS:SFHS samples of serum and urine or for targeted recruitment for other studies). They allow comparison and corroboration of the self-reported data from questionnaires and clinical measures collected during the study with SMR and prescribing data but also provide an opportunity to study endpoints and diseases not included in the original study.

Corresponding author email: pamela.linksted@ed.ac.uk

Navigating record linkage in Scotland and England and Wales: reviving the 6-Day Sample study

Brett, CE, Department of Psychology, University of Edinburgh
Deary, IJ, Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh

Health and wellbeing in old age is influenced by factors throughout the life course. Life course epidemiology provides an interdisciplinary framework for understanding how genetic and environmental factors such as socioeconomic status, occupational hazards and childhood deprivation influence health inequalities in old age. At present, few longitudinal studies offer information from childhood through to old age. The Scottish Mental Survey 1947 (SMS1947, n = 70,805, ~95% of the relevant population) has childhood intelligence data from individuals born in 1936 and attending schools in Scotland in June 1947. Representative subgroups of the SMS1947 provided additional sociological information in childhood. Those born on 6 days of 1936 (the 6-Day Sample, n = 1208) were followed up for 16 years to age 27. Their younger siblings also took an intelligence test and were followed up for several years.

The present study falls into two distinct parts. The first is a revival of the 6-Day Sample study involving tracing the 1208 Sample members and inviting survivors to a follow-up study. The second is a linkage between existing data on the SMS1947, its subgroups and younger siblings and health outcomes from the Scottish Morbidity Records, Health Episode Statistics and NHS Central Register. Both parts of the study involved linking in-hand data to records in Scotland and in England and Wales.

This presentation outlines the process of obtaining permission for these two aspects of the study: follow-up and data linkage. It highlights the organisations and procedures encountered and places these into the context of the shifting landscape of data linkage in Scotland and in England and Wales, and the differences between them. The SMS1947 offers an unusual opportunity to understand life course influences on health and wellbeing in old age in an entire year-of-birth population. Record linkage strengthens the value of the existing data and enables researchers to address some influences on health that stretch across several decades of people's lives. In addition to describing procedural aspects, an outline is given of the research plans and some initial findings. We also address the input that researchers can make to optimising record linkage systems.

Corresponding author email: caroline.brett@ed.ac.uk

Data linkage for pharmacovigilance using routine electronic health records

Kirby, B, Department of Child Health, University of Aberdeen
Simpson, C, eHealth Research Group, Centre for Population Health Science, University of Edinburgh
Helms, P, Department of Child Health, University of Aberdeen
Bennie, M, Information Services Division (ISD) Scotland, Edinburgh
Wood, R, Information Services Division (ISD) Scotland, Edinburgh
McLay, J, Department of Child Health, University of Aberdeen

Introduction

Despite the establishment of monitoring and pharmacovigilance systems, there is a recognised paucity of information specifically on the safety and effectiveness of medicines used in children. Within the constraints of safeguarding patient confidentiality and data protection, data linkage techniques offer real potential for linking routinely collected population based primary and secondary care datasets, using the Community Health Index (CHI) as a patient linkage key, in order to monitor the safety of new drugs and treatments; however the data first needs to be validated and its utility exemplified and that is the objective of this study.

Methods

Retrospective cohort studies, combined with data linkage techniques, will establish internal and external validity of the national Prescribing Information System (PIS) in Scotland, which is a whole population database that routinely collects all medicines dispensed in the community in Scotland. This study assesses the consistency of unique patient identifiers; measures the completeness and accuracy of the data; and examines the extent to which well established associations between drugs and adverse events can be reproduced using routine data.

Results

The CHI is a useful linking variable but has varying levels of completeness amongst datasets. On routine prescribing data a CHI number exists on over 90% of dispensed items since January 2010, with current levels reaching nearly 95%. Insulin prescriptions were identified for 96% of hospitalised type 1 diabetics aged under 20 years, and the rates of newly prescribed insulin for patients aged 0-30 were concordant with rates published in both Scottish and non-Scottish populations for T1 diabetes diagnoses. Asthma prescribing in children was found to be both complete (sensitivity 0.96 (95% CI 0.95-0.98)) and accurate (PPV 0.87 (95% CI 0.83-0.9)) when compared against a gold standard patient registry containing self-reported medicine use in the NHS Grampian region; asthma status was identified for 86% of patients who failed to respond to a postal survey. Finally, patients newly prescribed NSAID therapy were found to be around 2 to 4 times more likely to experience first time hospitalisation for gastrointestinal ulcers than those not exposed; the risks increased with age and concurrent use of antiplatelet and anticoagulant therapy and are concordant with those published in the medical literature.

Conclusion

The combination of results suggests that routine prescribing data in Scotland is consistent, complete and accurate; however several key variables were missing from the data that would have been useful for pharmacovigilance studies, such as indication, dose and frequency.

Corresponding author email: bradley.kirby@nhs.net

Use of record linkage for large-scale, epidemiological research: experience of UK Biobank

Sudlow, CLM, University of Edinburgh and UK Biobank
Allen, N, University of Oxford and UK Biobank
Adamska, L, University of Oxford and UK Biobank
Allan, V, University of Oxford and UK Biobank
Flaig, R, University of Edinburgh and UK Biobank

UK Biobank is a population-based, prospective study, providing open access to data from 500,000 men and women aged 40-69 years recruited in England, Scotland and Wales from 2006-2010. It aims to enable researchers to investigate the genetic and environmental determinants of a wide range of diseases of middle and old age. From its inception, the priority has been to facilitate extensive and precise measurement of exposures, along with detailed and rigorous follow-up for a wide range of health-related outcomes, and to promote innovative science by maximising access to the data.

A variety of different sources are being used to ascertain death, disease occurrence and other health-related information among participants during long-term follow-up. Several of these have an established track record in UK-based prospective epidemiological studies (e.g., death and cancer registries), whereas others, such as hospital episode records, have been used somewhat less widely. UK Biobank also aims to be among the first very large-scale prospective epidemiological studies in the UK to obtain linked data for all participants from electronic primary care records. In addition, we are investigating the potential benefits and practicalities of linking to a range of other electronic health-related data sources, including disease-specific registries, pharmacy records, dental records, screening programmes, private medical records, and databases related to social services (e.g., disability/incapacity benefit data), education (e.g., adult learning) and environment (e.g. household-level data for road traffic density and specific pollutants).

Record linkage on such a large scale requires reliable mechanisms to keep track of individual participants' health records across different data providers in England, Wales and Scotland and over time, whilst maintaining data security and protecting participants' identities. UK Biobank's linkages depend on successful partnerships with NHS data providers in all three countries, expert advice from our follow-up and outcomes working group, and ethical and regulatory approvals to hold key identifiers for linkage, including participants' National Health Service (NHS) numbers, and Community Health Index (CHI) numbers for participants in Scotland. The processing, cleaning, standardisation, presentation and interpretation of such data are complex and require a multidisciplinary approach involving epidemiological, clinical and technological expertise.

The availability of data from record linkages will allow researchers to investigate associations of UK Biobank's rich and increasingly extensive exposure data with health outcomes occurring during long-term follow-up, leading to improved understanding of the causes of conditions of major public health importance, including heart disease, cancers, stroke, arthritis, dementia and respiratory diseases.

Corresponding author email: cathie.sudlow@ed.ac.uk

The consultation on the establishment of a national health sector privacy advisory committee (NPAC) on the use of Scottish health records for research, statistical and related purposes.

Ruddy, P, NHS National Services Scotland

In the Scottish health service, health records are created in the context of a relationship of medical confidentiality between patients and the health professionals involved in their care. These professionals are employed by, or contracted to, Government-funded public authorities that are subject to a range of regulation. The number of stakeholders in the "good" governance of health records is therefore high, and certainly includes more than the pivotally important perspective of the patient.

Notwithstanding the potential challenges of meeting the requirements of these stakeholders, both the Scottish Government in its November 2012 strategy "Joined up data for better decisions: A strategy for improving data access and analysis" and the International Advisory Board of SHIP have indicated their wish to see the establishment of a national, Scottish multi-sector privacy advisory committee to oversee the use of data for research, analysis, statistical and related purposes.

In response to these calls, earlier in 2013, NHS National Services Scotland (NSS), a non-departmental public body answerable to the Scottish Government, began a consultation process on the setting up of a national privacy advisory committee (NPAC) to advise NHS Scotland on the release and linkage of actually and/or potentially identifiable personal health data held in Scotland for research, statistical and related purposes. Whilst the focus was to be on an NPAC for the Scottish healthcare sector, its implications and opportunities for the multi-sector NPAC, as envisaged by the Scottish Government and the SHIP International Advisory Board, were to be kept in mind.

Consultation included a number of stages, including a focus group of NHS Scotland data custodians and other key stakeholders, a formal consultation paper, and the consideration of an appropriate public engagement strategy. The practical and resource requirements and constraints of the setting up such a group - including those arising from the legal implications of the Data Protection Act 1998 and resource and budget constraints in financially challenging times for the Scottish public sector - were integral to considerations and discussions.

The results of the consultation will be presented. Whilst the success of an NPAC for the Scottish healthcare sector is outwith the control of NSS, the results will undoubtedly be informative to both the Scottish Government and SHIP in developing their vision of a multi-sector NPAC.

Corresponding author email: patricia.ruddy@nhs.net

Research Access to Health Administrative Data in Canada: Timelines, Processes, Successes and Bottlenecks

Meagher, NL, Population Data BC, University of British Columbia, Vancouver, BC, Canada
McGrail, K, Centre for Health Services and Policy Research, University of British Columbia, Vancouver, BC, Canada

How long does it take to get research access to health administrative data by jurisdiction in Canada? What is the process involved? These were the fundamental questions in a 2012 survey done of entities that handle research access to health administrative data in each of the 10 provinces and 3 territories.

Access times vary considerably, and while numbers reported will be given, the study illustrates the inherent challenge in standardizing reporting timelines. To unpack timelines it is important to understand the different models in place and their mechanisms for access. Even within a single country, this survey revealed quite different approaches to defining who can make requests for access to data, the requirements involved in making access requests, and what steps are involved in adjudicating and ultimately granting approval. Through this we gain some understanding of how each jurisdiction both defines and addresses privacy concerns.

One of the most significant findings is that there are many different approaches to the adjudication process. In all cases, research ethics boards were viewed as critical. Beyond that, there were very different processes and parties involved, from individual data stewards, to small groups to centralized (and separated) review committees. Researcher experience and data steward confidence in researchers' ability also were also raised as ongoing concerns.

While there is no one ideal model, there are opportunities to learn from both bottlenecks and innovations in each jurisdiction. This will be increasingly important as access to a wider variety of data sets from a broader array of agencies are of interest and available to researchers.

Corresponding author email: nancy.meagher@popdata.bc.ca

Cross-jurisdictional Data Linkage: Lessons from Australia

Smith, MB, Population Health Research Network, Australia

High quality, population-based research often requires analysis of individual-level data from a range of sources including across jurisdictional and national boundaries. Ethico-legal issues, custom and practice can prohibit or discourage this type of research.

Australia has one Federal, six state and two territory governments. Linkage of data across jurisdictional boundaries is particularly important in Australia because:

- Health and human services data is not held by a single level of government;
- The Australian population is mobile and there is significant movement across state/ territory boundaries; and
- There are large population centres that are adjacent to or straddle state boundaries

Australia has made significant progress over the last four years in overcoming barriers to cross-jurisdictional data linkage when consent is not sought.

In 2009, the Australian governments and academic partners commenced development of national, population-based data linkage infrastructure to support health research and related health system management. The development is being coordinated by a network called the Population Health Research Network (PHRN).

The infrastructure is focused on population health data collections but includes other human services data. The approach used is a distributed model which builds on well-developed existing infrastructure in two Australian jurisdictions (Western Australia and New South Wales/Australian Capital Territory). The new development involves building data linkage capability in the four states that did not have it (Queensland, Victoria, Tasmania and South Australia in collaboration with the Northern Territory) and for the Australian government, establishing units to conduct cross-jurisdictional data linkage i.e. linkage of data from two or more jurisdictions, and developing a secure data access facility. A two-stage process where data linkage is managed separately from the merging and provision of researcher access to linked content data is used.

There are significant challenges in linking data from different jurisdictions. Based on PHRN experience, these include:

- Differing legislative regimes;
- Variation in data linkage frameworks;
- Multiple approval processes and reporting requirements;
- Differences in underlying data and data management structures; and
- Competing demands for data custodians.

Despite the challenges, the PHRN has been able to achieve linkage of cross-jurisdictional data. This has required a collaborative environment, good governance mechanisms including consumer and community engagement, some funding to support development activities and plenty of persistence. The solutions being developed for cross-jurisdictional data linkage in Australia may be relevant in other countries as well as to cross-country linkage.

Corresponding author email: merrans@ichr.uwa.edu.au

Cross-country sharing of administrative health data: from aspirational to possible

Jorm, LR, Centre for Health Research, School of Medicine, University of Western Sydney, Campbelltown, Australia
Falster, MO, Centre for Health Research, School of Medicine, University of Western Sydney, Campbelltown, Australia
Khoo, J, The Sax Institute, Sydney, Australia
Churches, TR, The Sax Institute, Sydney, Australia

The scientific benefits of data sharing include accelerating the pace of discovery, promoting open inquiry, supporting diversity in analysis and interpretation and allowing results to be replicated and alternative hypotheses to be tested. In recognition of these, major international research funders, including the Australian National Health and Medical Research Council, UK Medical Research Council, UK Economic and Social Research Council, Wellcome Trust, Canadian Institutes of Health Research, and US National Institutes of Health are signatories to a Joint Statement on Data Sharing of Public Health Research [1], which sets out principles, goals and aspirations to promote the efficient use of research data to accelerate improvements in public health.

Data sharing presents particular benefits for researchers who use administrative health data, because of the potential for facilitating cross-sectoral research to support "joined-up" policymaking, and for enabling cross-country comparative studies to address shared challenges. However, health and medical researchers lag well behind those from other scientific domains including social, physical, computing and environmental sciences in their willingness to share their data, and their interest in using data generated by others [2]. In the case of researchers who use administrative health data, these attitudinal barriers are compounded by lack of clarity regarding the ownership of these data, and the ethical and legal conditions under which they might be shared. Indeed, the Joint Statement on Data Sharing of Public Health Research explicitly excludes routinely maintained databases to which the signatories contribute no funding. Although declaring that "Researchers creating datasets for secondary analysis from shared primary data are expected to share those datasets" [1], the Joint Statement is silent on the issue of secondary datasets generated from administrative health data.

Distributed methods for data synthesis without physical aggregation of data (such as DataShield [3]) and purpose-built secure remote data laboratory facilities (such as SURE [4]) present new possibilities for sharing administrative health data safely. However, practical experience in how to make this happen across institutions, let alone across countries, is thin on the ground. This presentation will describe the steps taken, barriers encountered, and progress made, in enabling researchers in Scotland to share Australian linked administrative health data for a collaborative project that is investigating preventable hospitalisations [5].

1. Walport M, Brest P. Sharing research data to improve public health. Lancet 2011; 377: 537-539.
2. Tenopir C, Allard S, Douglass K, et al. Data Sharing by Scientists: Practices and Perceptions. PLoS ONE 2011; 6(6): e21101.
3. Wolfson M, Wallace SE, Masca N, et al. DataSHIELD: resolving a conflict in contemporary bioscience e performing a pooled analysis of individual-level data without sharing the data. Int J Epidemiol 2010; 39: 1372e82.
4. About SURE. https://www.sure.org.au/about-sure
5. Jorm LR, Leyland AH, Blyth FM, et al. Assessing Preventable Hospitalisation InDicators (APHID): protocol for a data linkage study using cohort study and administrative data. BMJ Open 2012; 2: e002344.

Corresponding author email: l.jorm@uws.edu.a

Socioeconomic disadvantage, parental mental illness and deliberate self-harm in adolescents: a nested case-control study using record linkage

Hu, N, Telethon Institute for Child Health Research, Centre for Child Health Research, The University of Western Australia, WA, Australia; School of Population Health, The University of Western Australia, WA, Australia
Glauert, R, Telethon Institute for Child Health Research, Centre for Child Health Research, The University of Western Australia, WA, Australia
Li, J, Social Science Research Center, Berlin, Germany; Telethon Institute for Child Health Research, Centre for Child Health Research, The University of Western Australia, WA, Australia
Taylor, C, Telethon Institute for Child Health Research, Centre for Child Health Research, The University of Western Australia, WA, Australia

BACKGROUND Evidence has shown that restricted foetal growth, socioeconomic disadvantage, and parental psychiatric disorder and self-harm behaviour are all associated with an increased risk of deliberate self-harm (DSH) in adolescents. However, we know little about the extent to which these factors may interact and confound one another in shaping the risk of DSH behaviour of adolescents which results in hospital presentation.

OBJECTIVES To investigate how restricted foetal growth, familial socioeconomic disadvantage, and parental psychiatric and DSH presentation to hospital may interact and combine to influence the risk of DSH related hospital presentation among adolescents

METHODS A nested case-control sample was compiled from a birth cohort using administrative health records collected by the Western Australian (WA) government. Until 2011, 2,142 people with DSH related hospital presentations at the age between 10 and 20 were identified among people born during 1991-1999 in WA, and 42,840 controls were matched using incidence density sampling method by gender, birth year, and the date of the first DSH hospital presentation of the cases. Data were analysed with conditional logistic regression.

RESULTS Multivariate analysis showed that socioeconomic disadvantages in early life (measured by living in more disadvantaged neighbourhood SES, born to teenage parents, born to an unmarried mother, and high parity, adjusted OR ranging from 1.1-1.6) and parental psychiatric and DSH presentation (adjusted OR ranging from 1.7-6.0) independently increased the risk of DSH related hospital presentation of adolescents. These factors also partially accounted for the effects of restricted foetal growth (low birth weight and preterm/overdue birth) on the outcome. Further analysis showed that parental presentation in relation to different issues (psychiatric or DSH), presentation involved with different parent (father or mother), and parental presentation that occurred in different life stage of children (pregnancy, infancy, childhood, adolescence) were independently associated with a heightened risk of DSH of adolescents. These effects related to parental hospital presentation indicated specific trend patterns and were modified by the gender of adolescents.

CONCLUSIONS Familial and neighbourhood socioeconomic disadvantages and the history of parental psychiatric disorders and DSH have significant and independent impacts on the risk of DSH related hospital presentation of adolescents. The findings provide more insights into the mechanisms of how social and biological determinants interact to shape the risk of more severe DSH behaviour in adolescents that requires more critical medical attentions.

Corresponding author email: hnnathan@gmail.com

Mental Health Service Utilization Among High Risk Youth

Blackadar, R, Alberta Center for Child, Family and Community Research
Cui, X, Alberta Center for Child, Family and Community Research
Twilley, L, Alberta Center for Child, Family and Community Research
Werk, C, Alberta Center for Child, Family and Community Research
Bukutu, C, Alberta Center for Child, Family and Community Research
Lamba, J, Alberta Center for Child, Family and Community Research

Good mental health is a cornerstone of overall health. Personal and societal costs of mental health issues in youth are high. According to Scott et al (2001), youth with untreated mental health disorders have been found to have elevated rates of use of services such as health care, justice and corrections, special education programs, foster care, and income support.

Youth who report lower levels of satisfaction with and control over their lives, reduced sense of belonging, or impaired relationships have a higher rate of mental health problems (PHAC 2011). Aboriginal youth living off reserves are less likely than the overall population of youth to report having good or excellent mental health. Youth living in low income situations (who are more likely than the overall population to be Aboriginal, immigrant, or homeless youth) are more stressed than youth not living in low income situations. An increased understanding of mental health problems among these high risk populations can help policy makers develop appropriate policy, program and services to improve the outcomes of these youth.

The Alberta Center for Child, Family and Community Research has partnered with six Alberta provincial ministries to establish the Child and Youth Data Lab where administrative data is linked across these ministries such as Health, Human Services, Education and Justice and Solicitor General. Studies using the linked data can generate relevant research evidence for policy and practice. One of the studies looks at the service use pattern of Albertan youth with a focus on mental health, social economic status and education. Select results from the study are presented in the current submission.

In this study, data from five provincial government ministries, ten program areas were linked and 66,792 Albertan youth (12 to 24 years) who received services for mental health conditions in 2008/09 were examined. Approximately 11% of Albertan youth received services for mental health conditions. Higher mental health service uses were observed among youth who received maltreatment-related intervention services, were charged with an offence, were involved in corrections, or received income support. The rates were especially higher among Aboriginal youth in these populations. Possible interpretations and policy implications are discussed.

References

Scott, S., Knapp, M., Henderson, J., & Maughan, B. (2001). Financial cost of social exclusion: follow up study of antisocial children into adulthood. British Medical Journal, 323, 191-194.

Public Health Agency of Canada (2011). The Chief Public Health Officer's Report on the State of Public Health in Canada, 2011: Youth and Young Adults - Life in Transition. PHAC: Ottawa.

Corresponding author email: RBlackadar@research4children.com

The likelihood of a child developing autism spectrum disorder, intellectual disability or both is related to a mother's mental health status in the years before the birth

Fairthorne, J, Telethon Institute for Child Health Research, Centre for Child Health Research, University of Western Australia
Hammond, G, Telethon Institute for Child Health Research, Centre for Child Health Research, University of Western Australia
de Klerk, N, Telethon Institute for Child Health Research, Centre for Child Health Research, University of Western Australia
Bourke, J, Telethon Institute for Child Health Research, Centre for Child Health Research, University of Western Australia
Leonard, H, Telethon Institute for Child Health Research, Centre for Child Health Research, University of Western Australia

Aim To investigate the relationship between women's out-patient mental health contacts and the likelihood of subsequent offspring being diagnosed with intellectual disability, autism spectrum disorder (autism) or both.

Methods Data from the Midwives Notification System, the Mental Health Information System, and the Western Australian (WA) Intellectual Disability Database were linked using a unique code. The derived dataset included information on mental health (MH) outpatient contacts, maternal socio-demographic and birth details and subsequent autism/ID diagnoses in the offspring.

There were 213,656 WA mothers with children born from 1983-1999. Of these, 7,219 (3.4%) had a child diagnosed with ID, autism or both and 25,583 (12.0%) had had a MH outpatient contact. Mothers were grouped according to the disability of their eldest or "index" child as follows: "mild or moderate ID (mild ID)", "severe or profound ID (severe ID)", "Down syndrome", "other biomedical ID", "autism with ID" and "autism without ID". The comparator group was mothers of children with no ID or autism where the index child was the eldest child. The relationship between the mother having an outpatient contact and the disability or otherwise of her child was investigated using multinomial logistic regression controlling for parity, socio-economic status and mother's age, all at index birth.

Contacts with MH out-patients were grouped using ICD codes into seven diagnostic groups. The relationship between the rate of contact for a diagnostic group before the birth of the index child on later disability was examined using multinomial logistic regression, controlling for the same confounders.

Results Women with an outpatient contact before the index birth were more likely to have a child with mild ID [OR 2.23(1.99, 2.50)], autism with ID [OR 1.83(1.37, 2.44)], autism without ID [OR 2.56(1.85, 3.55) or "other biomedical ID" [2.19 (1.65, 2.91)]. Previous contacts for developmental disorders increased the odds of having a child with mild ID by 2.09 per contact/ year with 95% CI (1.33, 3.28) and the odds of autism with ID by 1.77 per contact/ year with 95% CI (1.02, 3.08). Previous contacts for childhood and adolescent disorders also increased odds of having a child with mild ID by 1.21 per contact/ year with 95% CI (1.11, 1.32).

Conclusion The likelihood of giving birth to a child with ID or autism is increased for women with MH contacts prior to their child's birth. Further detail and possible causes of these associations are discussed.

Corresponding author email: jfairthorne@ichr.uwa.edu.au

Prevalence of children's mental health disorders in survey data compared to population data: a comparison of two prospective cohorts

Wong, J, Telethon Institute for Child Health Research, The University of Western Australia
O'Donnell, M, Telethon Institute for Child Health Research
Glauert, R, Telethon Institute for Child Health Research
Bayliss, D, The University of Western Australia
Fletcher, J, The University of Western Australia

Prevalence estimates of childhood and adolescent mental health disorders appear to vary between 20 to 30% worldwide. It is therefore unsurprising that studies have yielded inconsistent findings in regards to the trends of prevalence of mental health disorders. Some reasons for the discrepancy in findings include the use of survey data and its associated attrition and selection bias. As population data may be used to overcome the issues presented by survey data, we compared the prevalence estimated from a prospective survey cohort (the Western Australian Pregnancy Cohort (Raine study)) to another estimate from a prospective population cohort (linked population data; Hospital Morbidity and Mental Health Registrations). As expected, the Raine cohort yielded a larger estimate of mental health prevalence when compared to the linked population data. However each cohort also revealed opposite trends of prevalence, where the Raine cohort showed the prevalence of mental health disorders to decrease as children age. To investigate whether attrition may have contributed to the decreasing trend of mental health prevalence in the survey cohort, the survey and population cohorts were linked to examine the characteristics of those who dropped out of the survey study. A logistic regression analysis showed that attrition was characterised by an increased number of risk factors, and their likelihood of accessing mental health services (as indicated by the linked population data) was different to those who remained in the survey study. We therefore recommend that estimates of prevalence be interpreted with the type of cohort in mind, as estimates from survey cohorts will provide different information to that from population cohorts.

Corresponding author email: wongj04@student.uwa.edu.au

International comparison of drugs users' cause-specific mortality needs to take account of epoch, demography, and injecting

Bird, SM, MRC Biostatistics Unit, Cambridge CB2 0SR
Pierce, M, National Drug Evidence Centre, University of Manchester, Ellen Wilkinson Building, M13 9PL
Merrall, ELC, Novartis Vaccines and Diagnostics, Amsterdam
Hutchinson , SJ, Health Protection Scotland, Glasgow G3 7LN
Hickman, M, School of Social and Community Medicine, University of Bristol, BS8 2PS
Millar, T, National Drug Evidence Centre, University of Manchester, Ellen Wilkinson Building, M13 9PL

Using record-linkage to deaths of two 'virtual' cohorts of UK drug users, we compare crude mortality rates and Standardized Mortality Ratios (C/SMRs) for specific causes of death: in the follow-up era 2001/02 to 2005/06 for nearly 70,000 drug treatment clients in the Scottish Drug Misuse Database, 1996-2006 (1,813 deaths in 256,309 person-years (pys); and for over 207,000 opiate or crack-cocaine users in the Drug Data Warehouse (DDW-OCC) derived from treatment or criminal justice sources in 2005/06 to 2008/09 in England and Wales. To account for late registration of deaths in E&W, information on deaths in the DDW-OCC cohort by 31 March 2009 (4,048 deaths in 571,646 pys) relates to deaths registered by 30 September 2011.

Methods: For SDMD cohort, individuals' risk period began at the date of the earliest SDMD registration on or after 1 April 2001, or at 1 April 2001 if previously SDMD-registered; for DDW-OCC cohort, at the date of earliest DDW observation on or after 1 April 2005, and at 1 April 2005 if already in treatment on 1 April 2005. Cause-specific C/SMRs were calculated at the ICD-10 chapter level, with observed deaths (O) compared to gender and age-appropriate national expected mortality (E) to derive SMRs (O/E). We applied the UK harmonized definition of drug-related deaths (DRDs, ICD-10 codes: F11-16; F18-19; X40-44; X60-64; X85; Y10-14). To avoid double-counting, DRDs were excluded from the ICD10 chapters of 'external causes' and 'mental and behavioural disorders'.

Findings: Rates (95% CI) per 10,000 pys were higher in the SDMD vs DDW-OCC cohort for: DRDs [36 (33-38) vs 31 (30-33) based on 915 & 1,797 DRDs); suicides [6.6 (5.3-7.3) vs 3.6 (3.2-4.2) based on 160 & 208 non-DRD suicides] and homicides [3.4 (2.7-4.1) vs 1.4 (1.1-1.8) based on 86 & 82 homicides]. By contrast, rates and SMRs (not shown) were lower in the SDMD vs DDW-OCC cohort for: diseases of the digestive system [5.7 (4.8-6.7) vs 7.1 (6.4-7.9) based on 147 & 408 deaths), circulatory system [4.1 (3.4-5.0) vs 7.2 (6.5-7.9) based on 105 & 411 deaths], cancer [2.3 (1.8-3.0) vs 5.4 (4.9-6.1) based on 60 & 311 deaths] and respiratory system [2.1 (1.6-2.8) vs 4.3 (3.8-4.9) based on 55 & 245 deaths]. Infectious disease rates were similar between cohorts (based on 62 & 156 deaths).

Interpretation: SDMD clients contributed only 28% of their pys at 35+ years of age compared to 45% for DDW-OCC clients. Other notable differences between the cohorts include that, at first registration, 60% of SDMD's opiate users reported having ever injected whereas declared ever injecting accounted for only 37% of pys for DDW's opiate-treatment sub-cohort of around 152,000 clients. Epoch, ageing and differently prevalent behavioural risks must all be borne in mind when attempting to compare cause-specific mortality rates between cohorts, such as for SDMD cohort in 2001-06 and DDW-OCC in 2005-09.

Corresponding author email: sheila.bird@mrc-bsu.cam.ac.uk

Smoking, Surgery, and Venous Thromboembolism Risk in Women: UK cohort linked to hospital admissions databases

Green, J, Cancer Epidemiology Unit, University of Oxford
Sweetland , S, Cancer Epidemiology Unit, University of Oxford
Parkin, L, University of Otago, New Zealand
Balkwill, A, Cancer Epidemiology Unit, University of Oxford
Reeves, G, Cancer Epidemiology Unit, University of Oxford
Beral, V, Cancer Epidemiology Unit, University of Oxford

Abstract based on published paper: Circulation 2013;127:1276-1282

Background
Evidence about the effect of smoking on venous thromboembolism risk, generally and in the postoperative period, is limited and inconsistent. We examined the incidence of venous thromboembolism in relation to smoking habits, both in the absence of surgery and in the first 12 postoperative weeks, in a large prospective study of women in the United Kingdom.

Methods
Over 1.3 million middle-aged UK women were recruited into the Million Women Study cohort between 1996 and 2001. Linkage to NHS registers for deaths, cancers and hospital admissions in England and Scotland provided virtually complete follow-up, and allowed exclusion of women with a prior history of cancer, recent surgery or venous thromboembolism at recruitment. Study questionnaire surveys provided detailed information on lifestyle exposures, including smoking. Cox proportional hazards models were used to obtain relative risks (RRs) with 95% confidence intervals (CIs) for incident venous thromboembolism in relation to smoking, taking potential confounding factors into account.

Results
During 6 years' follow-up of 1 162 718 women (mean age 56 years), 4630 were admitted to hospital for or died of venous thromboembolism. In the absence of surgery, current smokers had a significantly increased incidence of venous thromboembolism compared with never-smokers (adjusted relative risk 1.38, 95% confidence interval 1.28-1.48), with significantly greater risks in heavier than lighter smokers (relative risks 1.47 [95% confidence interval 1.34-1.62] and 1.29 [95% confidence interval 1.17-1.42] for ⩾15 versus <15 cigarettes per day). Current smokers were also more likely to have surgery than never-smokers (relative risk 1.12, 95% confidence interval 1.12-1.13). Among women who had surgery, the incidence of venous thromboembolism in the first 12 postoperative weeks was significantly greater in current than never-smokers (relative risk 1.16, 95% confidence interval 1.02-1.30).

Conclusions
Venous thromboembolism incidence was increased in current smokers, both in the absence of surgery and in the 12 weeks after surgery. Smoking is another factor to consider in the assessment of venous thromboembolism risk in patients undergoing surgery.

Corresponding author email: jane.green@ceu.ox.ac.uk

Change in alcohol outlet density and alcohol-related harm to population health (CHALICE)

Fone, D, Cardiff University; Dunstan, F, Cardiff University; White, J, Cardiff University; Webster, C, Cardiff University; Rodgers, S E, Swansea University; Lee, S, Cardiff University; Shiode, N, Cardiff University; Orford, S, Cardiff University; Weightman, A, Cardiff University; Brennan, I, University of Hull; Sivarajasingam, V, Cardiff University; Morgan, V, Cardiff University; Fry, R, Swansea University; Lyons, R, Swansea University

Excess alcohol consumption has serious adverse effects on health and violence-related harm. In the UK, around 37% of men and 29% of women drink to excess and 20% and 13% report binge drinking. The population health impact from a reduction in consumption is considerable. One proposed method to reduce consumption is to reduce availability through controls on alcohol outlets. In this study we investigate the impact of a change in outlet density on consumption and alcohol-related harm.

Cross-sectional evidence suggests that higher outlet density is associated with alcohol-related harm, particularly violence, but few longitudinal studies have investigated associations between outlet density and non-injury health outcomes. Most have not accounted for the effect of population migration on inequalities in alcohol-related outcomes. The population health impact from a reduction in consumption could be considerable. One proposed method to reduce consumption is to reduce availability through controls on alcohol outlets. In this study we investigate the impact of natural changes in outlet density on consumption and alcohol-related harm.

This paper describes the data linking methods employed and the initial results of a natural experiment to assess the effect of change in outlet density between 2005-2011, in Wales, UK; population 2.4 million aged 16 years and over. Data on outlets are held by the 22 local authorities in Wales under The Licensing Act 2003. Alcohol outlet densities have been computed in a Geographic Information System (GIS) as a function of network distance from household to outlet within 10 minute walking and driving buffer zones. This has been further enhanced by weighting the outlet by type, a variation on a two-step floating catchment methodology. The novel aspects of this analysis are the application of a GIS based network based method to estimate the association between changes in alcohol outlet density to alcohol-related harm in Wales, population. The study outcomes are change in (1) alcohol consumption using data from annual Welsh Health Surveys, (2) alcohol-related hospital admissions using the Patient Episode Database for Wales, (3) Accident & Emergency (A&E) department attendances between midnight-6am, and (4) alcohol-related violent crime against the person, using Police data.

The data have been anonymously linked within the Secure Anonymised Information Linkage (SAIL) Databank at individual and 2001 Census Lower Super Output Area levels. We will analyse the data using (1) longitudinal multilevel ordinal models of consumption and logistic models of hospital admissions and A&E attendance as a function of change in individual outlet exposure, adjusting for confounding variables, and (2) spatial models of the change in counts/rates of each outcome measure and outlet density. The impact on health inequalities will be assessed within deprivation strata and correction will be made for population migration using the Welsh Demographic Service.

This project was funded by the National Institute for Health Research Public Health Research (NIHR PHR) Programme (project number 09/3007/02).The views and opinions expressed therein are those of the authors and do not necessarily reflect those of the NIHR PHR Programme or the Department of Health.

Corresponding author email: r.j.fry@swansea.ac.uk

Using routine data to identify developmental problems: the Childhood Information for Learning and Development project

Thompson, L, University of Aberdeen
Marryat, L, University of Glasgow
Wood, R, NHS Information Services Division
Wilson, P, University of Aberdeen

Early experiences have a significant long term impact on health. Despite having relatively sophisticated routine data systems in Scotland, we do not routinely gather information on cognitive, social and emotional development, and it is not clear how child health data in general are collated and utilised.

The Childhood Information for Learning and Development (ChILD) project aimed to explore how existing (routinely collected) and novel (research) population based data on early childhood development can be used to support the improvement and evaluation of services for pre-school children. We conducted the study over three phases: (1) interview-based mapping of routine data systems across Scotland; (2) interview-based feedback from key stakeholders on Phase 1 results; (3) linkage of two research datasets (a - 30 month health visitor contact pilot; b - Strengths and Difficulties Questionnaire (SDQ) data at school entry) with routinely held child health data. We will give an overview of the outcomes of Phases 1 and 2 and then focus on the outcomes of Phase 3.

Phase 1 showed a large range of data is routinely gathered about children aged 5 and younger, but developmental measures are usually not recorded. Systems are universally available but do not always achieve universal coverage. More vulnerable children are more likely to be missed. Phase 2 showed that key stakeholders were impressed by the range of data available, but were not always confident in accessing and interpreting the data. Gaps in developmental information were identified. Phase 3 shows good feasibility in linking research data with routine data using probablistic and deterministic methods. Preliminary analyses suggest high-risk SDQ scores at age 5 are associated with mothers being young, smoking in pregnancy and living in a more deprived area, as well as health visitor assessed risk status at 6-8 weeks.

The ChILD project has shown that we gather a lot of data about young children, but not always of the right type or in a way accessible to those making decisions about children's services. Augmentation with research data is feasible and suggests that high SDQ scores at age 5 are associated with other markers of vulnerability. The potential for more integrated data systems should be explored.

Corresponding author email: Lucy.Thompson@abdn.ac.uk

Prediction of initiation and cessation of breast feeding from late pregnancy to 16 weeks: The Feeding Your Baby (FYB) cohort study

Donnan, PT, Dundee Epidemiology and Biostatistics Unit, University of Dundee
Dalzell, J, Directorate of Public Health, NHS Tayside
Symon, A, School of Nursing and Midwifery, University of Dundee
Rauchhaus, P, Dundee Epidemiology and Biostatistics Unit, University of Dundee
Monteith-Hodge, E, School of Nursing and Midwifery, University of Dundee
Kellett, G, School of Nursing and Midwifery, University of Dundee
Wyatt, JC, Leeds Institute of Health Sciences, University of Leeds
Whitford, HW, School of Nursing and Midwifery, University of Dundee

Introduction

Initiating and maintaining breastfeeding is relatively poor in the UK. In Scotland around two thirds initiate breastfeeding and by 6-8 weeks only one third or less are still breastfeeding. There are few data on continuous postnatal feeding and what psychological factors and attitudes are associated with initiating and stopping. The aims of the Feeding Your Baby (FYB) study were to collect real-time data on infant feeding using the novel validated method of SMS text messaging and consequently derive prediction models for both initiation and cessation of breastfeeding using demographic, psychological and obstetric variables.

Methods

The study design was a prospective cohort. Participants were pregnant women over 30 weeks gestation aged 16 years and above, living in Dundee, and booked to deliver at Ninewells Hospital, Dundee and able to speak English. Demographic data and psychological measures using a Theory of Planned Behaviour (TPB)-based questionnaire were obtained during pregnancy. Birth details, feeding method at birth and at hospital discharge were obtained from the Ninewells hospital database, Dundee, UK. Attitudes to breastfeeding were obtained through the Iowa Infant Feeding Assessment Scale (IIFAS). Breastfeeding women were followed-up by SMS text messages 2-weekly until 16 weeks or until breastfeeding was discontinued to ascertain feeding method and feeding intentions. The main outcome measures were initiation and cessation of breastfeeding.

Results

From the total cohort of women at delivery (n = 344) 68% (95% CI 63% to 73%) of women had started breastfeeding. Significant predictors of initiating breastfeeding were older age, parity, greater intention to breastfeed from TPB, higher IIFAS score as well as living with a husband or partner compared to living alone. For the final model the AUROC was 0.967. For those who initiated breastfeeding (n = 233), the strongest predictors of stopping were low intention to breastfeed from TPB, low IIFAS score and non-managerial / professional occupations. For stopping breastfeeding the discriminative ability or c-statistics were lower than for initiating with 0.65 for “exclusive” breastfeeding, and 0.69 for “any” breastfeeding respectively.

Conclusions

This work builds on our published study demonstrating the validity of SMS text messaging as a means of collecting data in a prospective study.1 The findings from this study will be used to inform the protocol for an intervention study to encourage and support prolonged breastfeeding as intentions appear to be a key intervention focus for initiation. The predictive models could be used to identify women at high risk of not initiating and also women at high risk of stopping for interventions to improve longevity of breastfeeding.

Reference

1. Whitford H, Donnan P, Symon A, Kellett G, Monteith-Hodge E, Rauchhaus P, et al. Evaluating the reliability, validity, acceptability and practicality of SMS text messaging as a tool to collect research data: results from the Feeding Your Baby project. Journal of the American Medical Informatics Association. amiajnl-2011-000785

Corresponding author email: p.t.donnan@dundee.ac.uk

Using linked data sets to inform the early years agenda: infant feeding and child health in Scotland

Ajetunmobi, O, Information Services Division (ISD), National Services Scotland
Whyte, B, Glasgow Centre for Population Health, Glasgow
Fleming , M, Information Services Division (ISD), National Services Scotland

Background: Breastfeeding is increasingly gaining prominence as a cost effective public health intervention to improve healthy child outcomes and reduce health inequalities, which are important priorities of the Scottish Government. The mode and duration of infant feeding has been associated with morbidity in early childhood. Socioeconomic and infant characteristics also influence the risk of infections and patterns of ill health in children; however these confounding factors are not often accounted for. Moreover, methodological issues such as sample size, data quality and varying definitions of breastfeeding have led to criticisms of many studies.

Objectives: To explore the impact of infant feeding patterns and trends on early child health hospitalisation and GP consultations in Scotland using a linkage of administrative data sets.

Methods: Administrative data comprising all registered births in Scotland between 1997 and 2009 (n = 731,611) were linked by the Information Services Division (ISD), National Services Scotland in a two phases. Phase 1 comprised the linkage of birth registration records (National Records of Scotland - NRS) to maternity records, infant health records, child health surveillance records and mortality records. In Phase 2, the linked dataset was extended to include hospital admission records; it also comprised primary care records for a subset of infants in the cohort (n=15,601). Descriptive and multivariate analyses (Cox regression and multilevel modelling) were conducted to quantify the risk of hospital admission and GP consultations for common childhood illnesses associated with the mode of infant feeding in the cohort followed up to March 2012.

Main Findings: A total of 137,905 infants with a valid 6 to 8 week review (27% of cohort with valid records) had been hospitalised over the study period. Bottle and mixed “bottle and breast” fed infants had higher rates of hospital admission (31% and 24% respectively) compared to exclusively breastfed infants (21%). After adjustment for a range of parental background, maternity and infant health characteristics, a greater relative risk of hospital admission was observed amongst bottle fed (HR: 1.23) and mixed fed infants (HR: 1.11). Similar patterns were observed in the subset of GP consultation records.

Conclusions: The study confirmed the benefits of exclusive breastfeeding to child health. It also highlighted other cultural, parental and maternity characteristics that influence infant feeding and child health within the Scottish context. The linked data set is a reliable resource for understanding infant feeding trends, which could be exploited in designing and targeting child health interventions in Scotland.

Corresponding author email: o.ajetunmobi@nhs.net

Linkage of Weather, Climate, and the Environment Data with Human Health and Wellbeing: MED-MI

Osborne, N, European Centre for Environment and Human Health, University of Exeter Medical School
Golding, B, UK Meteorological (Met) Office
Kessel, A, London School of Hygiene and Tropical Medicine
Depledge, M, European Centre for Environment and Human Health, University of Exeter Medical School
Cichowska, A, Public Health England
Bloomfield, D, University of Exeter
Hajat, S, London School of Hygiene and Tropical Medicine
Sabel, C, Department of Geography, University of Exeter
Haines, A, London School of Hygiene and Tropical Medicine
Fleming , L, European Centre for Environment and Human Health, University of Exeter Medical School

A large part of the global disease burden can be linked to environmental factors, underpinned by unhealthy behaviours. However, research into these linkages suffers from the lack of common tools and databases for carrying out investigations across many different scientific disciplines to explore these complex associations. The MED MI Partnership brings together leading organisations and researchers in climate, weather, environment, and human health and wellbeing.

The main aim is to create a central data and analysis source as an internet-based platform which will be a vital new common resource for medical and public health research in the UK and beyond. We will link and analyse complex meteorological, environmental, and epidemiological data. This is a vital step to translate this data and analysis resource into epidemiologic, clinical, and commercial collaborative applications, and thus, improved human health and wellbeing in a rapidly changing environment. Translational applications will:

- facilitate novel research into environmental exposures and health using integrated models;
- rapidly identify “hot spots” for targeted prevention, interventions, and research;
- provide healthcare practitioners, public health planners, and environmental managers with relevant information for improving services for locations and populations identified at risk;
- initiate and evaluate interventions to reduce the exposures, and thereby the health effects at both the individual and population levels;
- disseminate and provide access to data as part of outreach and engagement with the research community, policy makers and civil society.

Existing databases, currently stored in various locations/organizations, will be combined enabling climate, weather and environment data to be linked and analysed with human health and wellbeing data. With appropriate confidentiality and ethical safeguards, the Platform will be available to UK and other researchers

Corresponding author email: n.j.osborne@exeter.ac.uk

Address Cleaning and Temporal Linkage in Environmental Health Studies

Garwood, K, MRC-HPA Centre for Environment and Health, Imperial College London
Hambly, P, MRC-HPA Centre for Environment and Health, Imperial College London
Pearson, C, MRC-HPA Centre for Environment and Health, Imperial College London
Bickerstaffe, I, School of Social and Community Medicine, University of Bristol
Morris, T, School of Social and Community Medicine, University of Bristol
Northstone, K, School of Social and Community Medicine, University of Bristol
Hansell, A, MRC-HPA Centre for Environment and Health, Imperial College London
Gulliver, J, MRC-HPA Centre for Environment and Health, Imperial College London
De Hoogh, K, MRC-HPA Centre for Environment and Health, Imperial College London
Blangiardo, M, MRC-HPA Centre for Environment and Health, Imperial College London
Henderson, J, School of Social and Community Medicine, University of Bristol
Elliott, J, School of Social and Community Medicine, University of Bristol

There is growing interest in how long term exposure to various kinds of pollution influences different kinds of chronic disease. In environmental health studies it is often critically important to know the address history of the study population being modelled. One of the goals of these studies is to establish a contiguous address record for each case study member that covers a certain time frame. The address periods are linked with location-specified pollution data. The studies can then examine how exposure to specific pollutants relates to specific health outcomes such as respiratory and cardiovascular problems.

In one of our current study activities, we are assessing the relationship between exposure to air pollution and health problems in case study members who are part of the Avon Longitudinal Study of Parents and Children. Our analysis depends on establishing address histories which cover the first fifteen years of life. The location data needed for exposure data do not come from precise co-ordinates provided by GPS tracking systems. Instead, the locations are linked to the AddressPoint database for geocoding.

Relying on historical records from existing cohorts presents a major challenge: to re-purpose temporal address data that were originally gathered to support administrative rather than research ends. For many cohort groups, their whereabouts are tracked to support activities such as mail-outs or phone-based interviews. They are not tracked to monitor precisely their exposure to pollution sources. When address changes are recorded through electronic means, the data are often managed in simple spread sheets or database tables. The address records are often prone to various kinds of data entry errors that lead to gaps, overlaps and missing values in the address records.

We have developed an algorithm which cleans temporal address data based on assumptions that have been made about data entry habits. The algorithm is able to provide adjusted start and end dates for address periods and retain information about how they were changed from the original values. The data can be used to support sensitivity studies that compare the results of an analysis using all address periods and an analysis that uses those which have experienced limited or no change. As well as supporting a research activity, the algorithm has had to consider aspects of information governance.

Corresponding author email: kgarwood@imperial.ac.uk

The association between ambient modelled air pollution and birth outcomes in Scotland

Clemens, T, University of St Andrews
Dibben, C, University of St Andrews

A growing number of studies have reported a relationship between ambient air pollution and adverse birth outcomes. In this paper we examine the association between fetal development, prematurity and ambient background concentrations of sulphur dioxide (SO2), particulates up to 10µm in diameter (PM10) and Nitrogen Dioxide (NO2) at the mother's area of residence and place of work. We linked data from the Scottish Longitudinal Study (5% sample of the Scottish census in 1991 linked to 2001 census records) to maternity data from the Scottish morbidity record to identify a sample of singleton live births across Scotland. Modelled pollution data at a 1x1 km spatial resolution was obtained for the years 1994 to 2008 and linked to the census and birth records via the mothers residential and workplace postcode using a geographical information system. The association between pollution and mean birthweight, low birthweight < 2500g, small for gestational age, and prematurity was estimated adjusting for known confounders including ethnicity and smoking. The findings from the study will be presented and discussed together with a number of methodological matters arising from the study.

Corresponding author email: cjld@st-andrews.ac.uk

Missing data and multiple imputation in electronic health records

Bartlett, J, London School of Hygiene and Tropical Medicine
Petersen, I, UCL
Welch, C, UCL

Electronic health data have the potential to answer many important clinical and epidemiological questions. However, a number of methodological challenges hinder their full exploitation. Amongst these are the fact that by their very nature, routinely collected datasets predominantly consist of measurements and observations which were made as part of patients' clinical care. Consequently, only a small and likely unrepresentative subset of patients may have a “full” set of measurements for a particular analysis of interest. Analysing this subset will thus give imprecise, and potentially biased estimates. By defining a set of measurements which we would ideally like to have available for patients to perform our analysis, we can treat the unavailability of measurements as a missing data problem, and appeal to modern statistical techniques developed for handling missing data.

In this talk we will introduce the concept of a missing data mechanism, assumptions about which should guide us in how we tackle the problem of missing data. We will give an intuitive description Rubin’s taxonomy of missing completely at random (MCAR), missing at random (MAR) and missing not at random (MNAR), and explain under what conditions a complete case analysis gives valid inferences.

In analyses of routinely collected datasets, even in settings where a complete case analysis is unbiased, estimates are typically extremely inefficient because the observed data in the incomplete cases (usually the majority) is completely discarded.

An increasingly popular alternative is the method of multiple imputation (MI). We will describe the key steps involved in creating and analysing multiply imputed datasets, and explain the assumptions underlying MI. We will then briefly introduce the full conditional specification algorithm for MI, which has become an extremely popular approach to MI in complex datasets.

Corresponding author email: Jonathan.bartlett@lshtm.ac.uk

Longitudinal electronic health records and multiple imputation of missing data

Petersen, I, UCL
Bartlett, J, London School of Hygiene and Tropical Medicine
Welch, C, UCL

Often medical research is designed such that information for the regression analyses on covariates only comes from the "baseline" measurement of these variables. However, with longitudinal cohorts and electronic records there are now opportunities for more advanced study designs and analyses taking into account the change in values over time in these covariates. However, the methods for handling missing longitudinal data need to be compatible with the subsequent analyses of the data. It is possible to include several separate measurement points (time points) into one "standard" MI model. However, as we will discuss, using this imputation approach we may soon run into difficulties due to collinearity, overfitting and perfect predictions. The latter is particular a problem when dealing with categorical variables e.g. smoking status which are measured longitudinally. One way to overcome these issues may be to fit several MI models at separate time points and thus treat the longitudinal data as if they were a set of cross-sectional datasets. A modification of this approach is just to perform MI on baseline data, completely ignoring the available longitudinal information between baseline and outcome. While this approach may provide an immediate solution based on the existing implementation of MI it is clear that it does not take the advantage of the (typically correlated and therefore useful) longitudinal information. Further, information from earlier or later time points is not utilised, and imputed values of missing data at one time point may contradict or be inconsistent with earlier or later values. The new two-fold FCS MI algorithm addresses some of these issues, by only conditioning on measurements which are local in time.

In this talk we illustrate how the two-fold FCS MI algorithm works in practice and maximises the use of data available, even in situations where measurements are only made on a relatively small proportion of subjects at each time point. We discuss some of the strengths and limitations of the two-fold FCS MI algorithm, and contrast it with existing approaches to imputing longitudinal data. Lastly, we present results demonstrating the potential for efficiency gain through use of the two-fold approach compared to a more conventional 'baseline MI' approach.

Corresponding author email: i.petersen@ucl.ac.uk

The two-fold fully conditional specification algorithm

Welch, C, UCL
Petersen, I, UCL
Bartlett, J, London School of Hygiene and Tropical Medicine

Electronic health records often hold longitudinal records and many health indicators such as weight, systolic blood pressure and cholesterol are time-dependent and recorded intermittently with differed lengths of time between each measurement.

Using standard implementations of multiple imputation may be difficult in these settings, so in this talk we present a new approach to the multiple imputation of longitudinal electronic health records, the two-fold FCS algorithm, which is a modification of the fully conditional specification (FCS) approach to MI which was introduced in the preceding session.

In the application of the two-fold FCS algorithm we divide time into equal size time blocks. The algorithm then imputes missing values in the longitudinal data, imputing one time block, and then the next. The defining characteristic is that when imputing missing values at a particular time block, only measurements at that time block or adjacent time blocks are used. This obviates some of the principal difficulties which are typically encountered when attempting to apply a standard MI approach to imputing such longitudinal data. The rationale for the simplification is that measurements of health indicators at time blocks earlier or after the time block with imputed measurements are unlikely to provide substantial additional information than measurements at immediately adjacent time blocks.

In this talk we describe how the two-fold MI algorithm works, and how we have implemented the algorithm in freely available Stata program. We then illustrate the application of the two-fold program to an illustrative example based on a database of routinely collected health data.

Corresponding author email: catherine.welch@ucl.ac.uk

A model to reduce the barriers to execution of arbitrary code on potentially identifiable data without compromising privacy

Hubbard, TJP, Wellcome Trust Sanger Institute

The need to maintain the privacy of identifiable data frequently creates barriers to research, particularly to the application of arbitrary analysis code. Current solutions rely on either anonymisation and distribution of datasets, which depend on complex mechanisms to approve distribution and use, or the execution of analysis within a safe setting with either a restricted set of tools or an approval process for analysis code. In either case this limits the immediacy of testing a new hypothesis with a new algorithm on a new dataset that researchers have been used to an open data world. Being able to apply rapidly developing and complex algorithms has been particularly important to progress in the analysis of large genomics datasets.

Here an alternative approach is proposed by allowing arbitrary code to be executed over source data within a safe setting but where privacy is protected by limiting researchers to only being able to view summary data exported from the secure environment. Summaries would only be output via an API which places limits on the type and precision of output to prevent disclosure of identifiable information about any individuals. In order to lower the barriers to access code could be submitted to the safe setting as virtual machines based on templates provided by the safe setting.

This proposed scheme would be complementary rather than a substitute for existing systems. A limitation of this approach is that by preventing researchers from viewing source data it prevents data cleaning or investigations as to why analysis is failing. However, it could provide a fast track to test initial speculative hypotheses which could then be followed up by applications to access to source dataset for more detailed analysis.

Corresponding author email: th@sanger.ac.uk

Community Health e-Lab: Mobilising an Ethical Framework for Community-serving Uses of Individual Health Records

Ainsworth, JD, Health eResearch Centre (HeRC), University of Manchester

The second Caldicott review of information governance echoes government calls for the NHS and adult social care services to pledge “to anonymise the data collected during the course of your care and treatment and use it to support research and improve care for others” and “to inform you of research studies in which you may be eligible to participate”. However, there are significant barriers that prevent the efficient and effective reuse of health records. Firstly, the data need to be made available, but for the data custodians the theoretical risk of identity disclosure and unclear direct benefit to the community from which the data was generated, leads to over-caution and reluctance to make data available. Where data has been made available, the second problem emerges. Researchers who wish to use the data lack the knowledge of the policies in place for data access and can face inconsistent interpretation and application of the resulting requirements. This again has recently been highlighted in the second Caldicott review. The third problem is that procedures required by data custodians to ensure effective information governance through the whole life-cycle of research (from data acquisition to archiving) are often opaque, and incomplete. Consequently the researchers lack the procedural knowledge to efficiently and effectively use the data.

To overcome these problems we have developed the eLab. An eLab is a secure environment for managing, exploring and analysing anonymised data from health records, accessible through a web browser. The eLab addresses these three information governance barriers. It embeds a re-identification service that reveals patient identities solely to professionals with a legitimate relationship to a patient. This enables health care organisations to provide screening, safety monitoring and research recruitment services, providing clear benefit to the population. By establishing each eLab as a NRES Research Database a local governance board enables consistent application of policy and considers ethics in the context of the community it serves. The eLab then supports this by binding the information governance to the data, by explicitly embedding it in the researchers workflow. Within the eLab environment the researcher can identify the variables required for their research questions, submit an application to the governance board, received the decision and finally access the data in one place. For the data custodians, a complete audit trail is automatically generated. The eLab model provides an integrated environment for health record reuse that benefits the researchers, the data custodians, and the population that create the data.

Corresponding author email: john.ainsworth@manchester.ac.uk

Secure Patient Data Integration for In Silico Oncology

Coveney, PV, UCL

Information arising from post-genomic research and combined genetic and clinical trials, along side advances in high-performance computing and informatics, are rapidly providing the medical and scientific community with an enormous opportunity to improve prognosis of patients with cancer by individualizing treatment and moving forward to personalized medicine. Multi-level data collection within clinico-genomic trials and interdisciplinary analysis by clinicians, molecular biologists and other specialists involved in life science is mandatory to further improve the outcome of cancer patients' treatment. It is essential to merge the research results of biomolecular findings, imaging studies, scientific literature and clinical data from patients and to enable users to easily link, analyse and share data. The EU FP7 funded p-medicine project aims to do this by bringing together clinical trials data from across Europe, and applying informatics techniques to the data to make personalised patient health forecasts which feed back into the clinical decision making process.

Health informatics relies, in part, upon computational simulation, modelling and data mining methods. These, in turn, rely upon information from multiple sources, which is currently organised without reference to universal standards of terminology, language or schema. At the heart of p-medicine is a distributed data warehousing system designed to colocate pseudonymised data coming from the project's clinical partners [1]. This warehouse, called ampoule-pi, is an open source/open standards based tool for securely storing and maintaining data from diverse sources integrated semantically to enable reporting and analysis. It meets the challenge of taking heterogeneous data from multiple hospitals across Europe and presenting it in a consistent and standardised manner to predictive computational tools.

The warehouse takes into account the information governance directives agreed by the project consortium, through a system of role based access control, which restricts the data that individual users are able to view and operate on. The security problems associated with storing and processing identifiable patient data dictate that we work with anonymised digital material. However, we cannot inform clinical decisions if we are unable to link the results of a computational study back to a specific patient. For this reason, the warehouse integrates with data pseudonymisation services provided by Custodix. These services eliminate identifiable patient fields before data leaves the hospital, replacing them with pseudonymous identifiers. A trusted third-party holds a lookup table that is able to map an identifier back to an individual, should it become necessary to invoke a computational result to inform a clinical decision.

[1] B. Jefferys, I. Nwankwo, E. Neri, D. Chang, L. Shamardin, S. H.nold, N. Graf, N. Forgo and P.V. Coveney, "Navigating legal constraints in clinical data warehousing: a case study in personalised medicine", J R Soc Interface Focus, 3 (2), 20120088, (2013).

Corresponding author email: p.v.coveney@ucl.ac.uk

Atrial fibrillation incidence and its hazard in the hypertensive population: a risk prediction function from and for clinical practice.

Alves-i-Cabratosa, L, Vascular Health Research Group. Institut d'Investigació en Atenció Primària. Research Support Unit.IDIAP Jordi Gol. Catalonia, Spain
Comas-Cufí, M, Research Support Unit. Girona. Institut d'Investigació en Atenció Primària. IDIAP Jordi Gol. Catalonia, Spain
Garcia-Gil, MM, System for the Development of Research in Primary Care. Institut d'Investigació en Atenció Primària. Research Support Unit.IDIAP Jordi Gol. Catalonia, Spain Et al.

Background
Atrial fibrillation (AF) stands out as the most common arrhythmia in clinical practice. Amongst its risk factors, hypertension is the principal one, due to its high prevalence. Estimation of the incidence of AF and its determinants in the hypertensive population remains elusive; but it could be improved with better networking of the different available registers.

Objective
To estimate AF incidence and to determine its predictors amongst patients with hypertension, without previous cardiovascular disease (CVD), within the SIDIAP (System for the Development of Research in Primary Care) and hospital discharge databases.

Methods
We conducted an historical cohort study between 1 July 2006 and 31 December 2011.
SIDIAP contains anonymised longitudinal individual patient information including sociodemographic characteristics, morbidity (ICD-10), clinical and lifestyle variables, laboratory tests, and pharmacy invoice data. Hospital discharge database contains anonymised information on discharge diagnosis (ICD-9) and procedures.
We included 266000 hypertensive patients aged 55 years or over at the time of study entry. We excluded individuals with previous diagnosis of coded AF or other CVD.
The outcome of our study was incident AF; for its estimation, SIDIAP was linked to the hospital discharges database. For risk prediction, a derivation and a validation cohort- 60% and 40% of the whole database, respectively- were defined, and a Cox proportional hazards model was fitted.

Results
We found an AF incidence of 10.61 per 1000 person-year (95% CI: 10.42-10.80).
In the preliminary estimation model, the following variables associated directly with incident AF ($p<0.01$): age, sex, alcoholism, chronic obstructive pulmonary disease, valvular heart disease, heart failure, obesity, high systolic and diastolic blood pressure values -over 180 and/or 110 mmhg respectively; the following variables associated inversely with incident AF: total cholesterol and glomerular filtration rate.
The Brier score on the validation model was 0.027 and the area under the receiver operating characteristic curve was 0.73 (95% CI 0.72-0.74) and 0.73 (95% CI 0.71- 0.74) on the derivation and validation cohort respectively.

Conclusions
The linkage of the SIDIAP and hospital discharge databases provides an accurate estimation of the incidence of AF, comparable to that found within the literature. Our preliminary model performs satisfactorily on AF prediction within the hypertensive population, using variables measured in routine clinical practice. Patients at high risk could then be selected for diagnostic and preventative measures.

Corresponding author email: lalves@idiapjgol.info

National Sexual Health (NaSH) IT System in Scotland: The potential for sexual health research

McDaid, LM, MRC/CSO Social and Public Health Sciences Unit, University of Glasgow
Docherty, S, University of Glasgow
Winter, AJ, Sandyford Sexual Health Services, NHS Greater Glasgow and Clyde

Background: Specialist sexual health settings present specific challenges to electronic patient record (EPR) systems, most important of which is patient desire for discretion and anonymity. Bespoke "stand-alone" sexual health EPR systems have been developed across the world, including the National Sexual Health (NaSH) system in Scotland, which went live in 2008. Here we discuss the key issues surrounding the potential secondary use of NaSH data for sexual health research.

Methods: A scoping review in three stages: policy review of NaSH documentation; review of EPR issues reported by an international selection of clinics known to be using computerized clinical systems; and review of more general methodological issues related to the use of EPR.

Results: NaSH entails a data set of over 700,000 registered patients. Data on more than 300,000 attendances are recorded annually. Data include medical, family and sexual history, reproductive health and contraception, social and lifestyle factors, test requests/results (>400,000 annually), patient actions/recalls, prescriptions, symptoms, physical examination details, partner notification, and referrals. NaSH allows patient-centred choice of an anonymous identifier or CHI number, which could facilitate record linkage, with up to 75% of patients agreeing to use CHI for registration. Key issues in the use of the data are: data collection and completeness; storage and retrieval; and research governance. An anonymised "data view" has been created and is in use nationally and regionally for business reporting. Not all NHS boards complete the agreed minimum data set for each patient and use of NaSH as a true paperless "real time" EPR has been problematic in some areas. The reporting database only reflects current, visible data, and while episode-based data remain true, lifetime sexuality and smoking status, for example, can change over time, and the "original" or preceding data are written over; precluding any longitudinal analysis without archiving data. Similarly, longer term retention of NaSH data and availability to researchers out with the NHS are issues that have yet to be addressed.

Conclusions: Interrogating anonymised NaSH data would enable researchers to make better use of existing sexual health data in Scotland, be cheaper than initiating large-scale surveys, and give access to high-risk populations, but has to to address conflict between the need for comprehensive and complete data for research and for a routine clinical system to function in a routine way. Concerns over data collection, storage and retention should be considered within the context of the wider public health and research benefit.

Corresponding author email: l.mcdaid@sphsu.mrc.ac.uk

Mortality in Scottish prisoners: a cohort study

Graham, L, Information Services Division, NHS National Services Scotland
Stockton, D, Information Services Division, NHS National Services Scotland
Fischbacher, C, Information Services Division, NHS National Services Scotland
Fraser, A, Scottish Prison Service
Fleming, M, Information Services Division, NHS National Services Scotland
Grieg, K, Information Services Division, NHS National Services Scotland

Background. Mortality is known to be increased among those who have been in prison but understanding of the causes of increased mortality is limited by the fact that the majority of deaths occur outside prison and are not recorded in prison databases. Relatively few studies have examined long term mortality outcomes among prisoners and all cause mortality has not been previously described for prisoners in Scotland.

Methods. Standard probabilistic record linkage methods were used to link the Scottish Prison Service database to routine death registration data for individuals imprisoned in Scotland for the first time between 1st January 1996 and 31st December 2007.

Findings. Among 76,627 individuals there were 4,414 deaths (3,982 in men) during a median follow up time of 6.9 years for men and 6.1 years for women. Compared to the general population the age-standardised mortality rate among prisoners was 3.3 (95% CI: 3.2, 3.4) times higher for men and 7.6 (6.9, 8.3) times higher for women. Further adjustment for an area measure of deprivation accounted for part of this excess (adjusted relative risks 2.3 (2.2, 2.4) and 5.7 (5.1, 6.2) for men and women respectively). The most common cause of death for both men and women was mental and behavioural disorders. Relative risks were highest for drug and alcohol related causes, suicide and homicide and were markedly higher among women than men. For example, compared to the general population the risk of suicide was 3.5 (3.3, 3.8) times higher in men and 11.7 (9.4, 14.5) time higher in women. In addition to the causes mentioned, significantly increased death rates were noted for cardiovascular disease, lung cancer, other respiratory disease, infectious causes, transport injuries and homicide. Out of prison death rates were highest in the first week after discharge from prison. Mortality rates were lower in those with longer total duration in prison and highest in those with multiple episodes in prison.

Interpretation. Linkage of data from health and criminal justice systems has the potential to increase understanding of mortality of a highly disadvantaged group. People who have been imprisoned in Scotland experience substantial excess mortality that is only partly explained by an area measure of deprivation. The association of increased mortality with multiple shorter periods in prison and the concentration of deaths in the early period after prison discharge have implications for policy and practice.

Corresponding author email: colin.fischbacher@nhs.net

Determination of school-based contextual factors and their association with the prevalence of overweight and obesity in children

Williams, AJ, University of Exeter Medical School and Children's Health and Exercise Research Centre, University of Exeter
Wyatt, KM, University of Exeter Medical School
Williams, CA, Children's Health and Exercise Research Centre, University of Exeter
Henley, WE, University of Exeter Medical School

The international prioritisation of the treatment and prevention of non-communicable diseases has highlighted the obesity epidemic. Given the complex nature of obesity, research into the causes, treatments and prevention strategies require multiple research methodologies. Within England and Wales the introduction of the National Child Measurement Programme (NCMP) has led to the creation of a database of repeated cross-sectional demographic and weight status data for children aged 4-5 (Reception) and 10-11 (Year 6) years. Using the data from this programme a study was undertaken to examine the association between the school attended and the pupils' weight status and to identify which school-based contextual factors were associated with overweight and obesity.

The NCMP database contains the body mass index standard deviation score (BMI-SDS), gender, age and ethnicity of each participating pupil alongside the lower super output area (LSOA) of the pupil and their school, and a unique school reference number (USRN). The individual and school LSOA were linked to indices of socioeconomic status. School contextual data were acquired from the regular county council school surveys, as well as from publicly accessible databases such as EduBase. These data were compiled into a schools data set and merged with the NCMP data set using the USRN. The final data set spanned 5 years (2006/07 - 2010/11) and contained 62,554 pupils from 319 schools across Devon, as well as, 40 potential explanatory variables which were organised into the following thematic clusters: demography, socioeconomic status, built environment, physical activity, diet and ethos.

Each of the five years of data were analysed separately, so that significant associations between explanatory variables and BMI-SDS could be compared across years. Within each year, four regression models were developed: single level, separate year group (Reception and Year 6) two-level (pupil > school), and three-level (pupil > year group > school) . Only the socioeconomic status of a school's location was consistently associated with higher pupil BMI-SDS after adjustment for the individual factors. It had been hypothesised that the school effect would be bigger in Year 6 than Reception due to the additional exposure to the school environment. However, the school intraclass correlation coefficients were larger in Reception (<4%) than Year 6 (≈ 1%) which suggests that this effect may not relate to schools. School adjusted mean BMI-SDS differed from the population mean by an average of <0.2 standard deviation which in Reception pupils equates to approximately 0.4 kg, and 1.3 kg in Year 6 pupils.

Corresponding author email: andrew.williams@pcmd.ac.uk

Body mass index and coronary heart disease in the Million Women Study: a prospective study

Canoy, D; Cairns, BJ; Balkwill, A; Wright, FL; Green, J; Reeves, G; Beral, V
Cancer Epidemiology Unit, University of Oxford, UK

Background: While coronary heart disease (CHD) mortality risk increases with increasing body mass index (BMI), low BMI may also be associated with an elevated risk. There is limited information on the pattern of the relation of CHD incidence and mortality risk across a wide range of BMI values. We examined the prospective relation between BMI and CHD and compared the shape of the association for incident and fatal outcomes of the disease in the Million Women Study.

Methods: Over 1.2 million women (mean age=56 years) who completed a health and lifestyle questionnaire and without heart disease, stroke, or cancer (except non-melanoma skin cancer) at baseline were followed prospectively for 9 years on average. Participants were linked with the National Health Service (NHS) Central Registers for information on deaths and the NHS databases for information on hospital admissions (using the Hospital Episode Statistics and the Scottish Morbidity Records for participants in England and Scotland, respectively) to identify women with a hospital admission diagnosis of CHD (ICD-10 I20 to I25) and women with CHD as the underlying cause of death. Adjusted relative risks and 20-year cumulative incidence rates from age 55-74 years were calculated for CHD using Cox regression.

Results: After excluding the first four years of follow-up, 32,465 women had a first coronary event (hospitalisation or death) during follow-up. The cumulative incidence of CHD from age 55 to 74 years increased progressively with increasing BMI, from 1 in 11 (95% confidence interval, 1 in 10-12) for a BMI of 20 kg/m2, to 1 in 6 (95% confidence interval, 1 in 5-7) for a BMI of 34 kg/m2. A 10-kg/m2 BMI increase conferred a similar risk to a 5-year increase in chronological age. By contrast to incident disease, the relation between BMI and CHD mortality (N=2431) was J-shaped. Relative risks were greater for CHD mortality than for incident disease in the lowest (<20 kg/m2) and highest (⩾ 35 kg/m2) BMI categories.

Conclusion: Coronary heart disease incidence in women increased progressively with increasing BMI. The shape of the relation with BMI differs for incident and fatal disease. The importance of capturing incident coronary events using hospital admission databases will be discussed within the context of answering research questions that may require statistical power of large-scale studies.

Reference: Canoy D, Cairns BJ, Balkwill A, Wright FL, Green J, Reeves G, Beral V. Body mass index and incident coronary heart disease in women: a population-based prospective study. BMC Med. 2013;11 (advanced online publication).

Corresponding author email: dexter.canoy@ceu.ox.ac.uk

SurgiCal Obesity Treatment Study: using record linkage for health technology appraisal

Logue, J, On behalf of the SCOTS Investigators Group, University of Glasgow
Stewart, S, University of Glasgow

As the prevalence of severe obesity has increased to epidemic levels, the provision of bariatric (weight loss) surgery in both the NHS and private sector has also increased. While bariatric surgery is thought to be an effective and safe treatment for severe obesity, like most novel surgical techniques the evidence base for long term outcomes and complications is scarce.

For this reason, the National Institute of Health Research advertised a commissioned call for a longitudinal study of bariatric surgery. The SurgiCal Obesity Treatment Study (SCOTS), run by the University of Glasgow, has been funded to follow-up 2000 patients undergoing bariatric across Scotland for 10 years post-operatively, with an eventual health economic evaluation planned. The majority of this follow-up is via record linkage to routine health records to record clinical outcomes including post-operative complications.

SCOTS has just started to recruit patients from all NHS and private hospitals within Scotland. Two IT systems have been developed, one for use by patients and one by surgical teams. Patients will complete a variety of questionnaires including quality of life and symptom specific questionnaires on an annual basis via the SCOTS website. Surgical teams will have access to a unique patient management system where they can record details of all interactions with their patients, including operative details, weight and gastric band adjustments.

There is planned linkage to a variety of data sources in NHS Scotland, including Sci-Diabetes, e-Phamacy, Sci-store, and Scottish Morbidity Records. This will allow long term outcomes to be ascertained including the effect on diabetes complications, cardiovascular disease, cancer, nutritional deficiencies and medication usage. It also allows detailed analysis of post-operative complications such as DVT (via ultrasound reports), post-op infections (from microbiology results and antibiotic prescription) and haemorrhage (from blood transfusion records). This wealth of information will mean that SCOTS will be the largest and most comprehensive study of bariatric surgery in the world.

The model that SCOTS follows could be easily replicated for a large variety of health technology appraisals, particularly for surgery and complex interventions that have not benefitted from in depth randomised controlled trials prior to implementation in the NHS. This provides an exciting opportunity for Scotland to capitalise on its record linkage potential at a time where there is growing recognition of the need for better evidence of efficacy and safety of new and existing treatments.

Corresponding author email: jennifer.logue@glasgow.ac.uk

Body mass index and risks of haemorrhagic and ischaemic stroke in women: UK cohort linked to routine hospital discharge data

Kroll, M E, Cancer Epidemiology Unit, University of Oxford, UK
Reeves, G K, Cancer Epidemiology Unit, University of Oxford, UK
Green, J, Cancer Epidemiology Unit, University of Oxford, UK
Beral, V, Cancer Epidemiology Unit, University of Oxford, UK
Sudlow, C, Division of Clinical Neurosciences, University of Edinburgh, UK

Background: Stroke is a major cause of disability and death worldwide. Although increasing body mass index (BMI) is a known risk factor for stroke in general, the evidence is weighted by findings for ischaemic stroke, the most common pathological type. Prospective studies to date have generally accrued too few stroke outcomes with information on pathological type to estimate reliably the separate associations of BMI with haemorrhagic as well as ischaemic stroke types.

Objectives: We aimed to clarify these associations by linking National Health Service central databases to individual participant data from a large UK cohort study.

Methods: The Million Women Study recruited 1.3 million middle-aged women in England and Scotland during 1996-2001, through the UK national breast-cancer screening program. Health and lifestyle factors at recruitment, including height and weight, were reported by questionnaire. Linkage to National Health Service Central Registers and hospital discharge data (Hospital Episode Statistics for England, and the Scottish Morbidity Record) provided virtually complete follow-up for deaths and hospital admissions, including date and cause coded to the 10th edition of the International Classification of Diseases. Incident stroke was defined as first hospital admission mentioning stroke, or death due to stroke. We estimated relative risk using Cox regression, adjusted for socioeconomic status, region, smoking, alcohol and physical activity.

Results: 13,754 incident strokes (5,994 ischaemic, 1914 intracerebral haemorrhage, 2440 subarachnoid haemorrhage, 3406 uncertain pathological type) occurred during an average follow-up of 9.2 years. Increasing BMI was associated with increased risk of ischaemic stroke (relative risk 1.21 [95% confidence interval 1.17-1.25] per 5 kg/m2 increase in BMI) but decreased risk of haemorrhagic stroke (intracerebral or subarachnoid, relative risk 0.90 [0.86-0.93] per 5 kg/m2 increase in BMI) (P<0.00001 for heterogeneity). There was no significant difference between intracerebral and subarachnoid haemorrhage in the association with BMI (P=0.4 for heterogeneity). In women who smoked or reported being treated for hypertension at recruitment, the increasing trend for ischaemic stroke was weaker, and the decreasing trend for haemorrhagic stroke was stronger.

Conclusions: In middle-aged UK women, increasing BMI was associated with increased risk of incident ischaemic stroke, but decreased risk of both intracerebral and subarachnoid haemorrhage.

Corresponding author email: mary.kroll@ceu.ox.ac.uk

Recurrent admissions in adolescents with victimisation-related injury: are the same adolescents also admitted for injury related to drug or alcohol misuse or self-harm?

Herbert, A, 1Centre for Paediatric Epidemiology and Biostatistics, Institute of Child Health, UCL, WC1N 1EH
Li, L, 1Centre for Paediatric Epidemiology and Biostatistics, Institute of Child Health, UCL, WC1N 1EH
Gonzalez-Izquierdo, A, 1Centre for Paediatric Epidemiology and Biostatistics, Institute of Child Health, UCL, WC1N 1EH
Gilbert, R, 1Centre for Paediatric Epidemiology and Biostatistics, Institute of Child Health, UCL, WC1N 1EH

Background & Objective: Community surveys and cohort studies indicate that adolescents exposed to victimisation have high risks of misusing drugs or alcohol or self-harming. We used linked hospital administrative data to determine the extent to which adolescents admitted for injury related to either victimisation, drug/alcohol misuse or self-harm are the same people.

Methods: We extracted hospitalisation trajectories for adolescents with an unplanned (acute) injury admission for adversity from 1997-2011. We defined "adversity" by clusters of ICD-10 codes for victimisation (e.g., assault, maltreatment) and substance misuse (e.g., drug/ alcohol misuse or intentional self-harm by poisoning). We determined the proportion of adolescents admitted for either victimization-related injury only, substance misuse-related injury only, or both types, during adolescence (10-19 years). We then compared the rates of unplanned re-admission for any reason, according to these groups.

Results: 168,088/777,337 (22%) adolescents with an unplanned injury admission were coded for any adversity during their adolescence. 28% (46,882/168,088) of these adolescents experienced both victimisation and substance misuse during this period, and 71% (33,555/46,882) of these experienced both at a simultaneous admission. 19% (31,896/168,088) and 53% (89,310/168,088) of adolescents with an adversity-related injury admission were admitted for injury related to victimisation only and substance misuse only, respectively. Rates of unplanned re-admission were higher for adolescents who had experienced both victimisation and substance misuse during adolescence than those who had experienced only one type (46% [21,692/46,882] for both, 27% [8,695/31,896] for victimization only, 37% [33,010/89,310] for substance misuse only and 19% [113,862/609,249] for no adversity).

Conclusions: A fifth of adolescents admitted to hospital suffer any type of adversity. Over a quarter of these experienced more than one type of adversity and had a higher unplanned hospital re-admission rate, compared with those with only one type of adversity or no adversity. Therefore, considering other current and previous adversity-related injury as well as current primary diagnosis for patient management may reduce the risk of unplanned re-admission, particularly in the case of victimisation. Analysing other hospital data such as accident and emergency visits, may indicate whether or not there are similar patterns in hospital attendances without admission.

Corresponding author email: annie.herbert.12@ucl.ac.uk

Modelling risk of smoking related disease linked to deprivation: comparison of two linked data sets.

Olajide, D, Health Economics Research Unit, University of Aberdeen
Ludbrook, A, Health Economics Research Unit, University of Aberdeen

Background

Linking disease risk associated with health behaviours to deprivation can assist in targeting interventions and addressing health inequalities. Relevant data sets available in Scotland are the Scottish Health Survey (SHeS) and the Scottish Longitudinal Survey (SLS), administratively linked to Scottish Morbidity Records (SMR). The SHeS includes self reported health behaviours, but is potentially limited by small numbers when it comes to investigating specific diseases. The SLS has large numbers but does not contain individual health behaviour data. However, small area estimates of smoking probabilities have been developed. This study was a novel attempt to demonstrate the potential of such area based statistics by comparing results from separate analyses of the two data sets.

Methods

Probit regression was used to estimate the probability of developing a smoking related disease controlling for a range of social, demographic and health variables. Each data set was analysed independently using the most comparable variable definitions available. The SHeS data were analysed with and without data on health behaviours other than smoking (alcohol, diet and physical activity).

The samples consisted of N=199,585(SLS) and N=20315 (SHeS) observations for individuals in the 2001 Census (SLS) or 1995, 1998 and 2003 SHeS, aged 16 and above, who consented to record linkage and were available for follow up. Both surveys included demographic variables; socioeconomic characteristics; reported health status; and previous health history. Smoking behaviour was measured by two indicators of area level probability of smoking (SLS) and by individual self-report (SHeS).

Results

Sample statistics were similar in demographic variables but exhibited some variation in the deprivation and health variables. The first incidence of a smoking related disease was slightly higher in the SHeS sample (23.8% v 22.7%) whereas the prevalence of pre-survey disease was lower (11.1% v 14.3%).

In terms of the sign and significance of coefficients in the probit models, there was broad agreement between the SLS and SHeS models. Disagreement on significance was generally due to insignificant coefficients in the SHeS model, not explained by the smaller sample size. Disagreement on sign, with both coefficients being significant, occurred in only 3 instances; age squared, married/cohabiting and low education. Predicted probabilities of smoking related diseases were estimated for specified cohorts to compare the performance of the models.

Conclusion

Further investigation is required to consider disease specific events and the value of adding small area measures of other health behaviours to the SLS.

Corresponding author email: d.olajide@abdn.ac.uk

Record linkage validation and behavioural-risk study: drugs-related deaths soon after hospital-discharge for drug treatment clients in Scotland, 1996-2010

White, S, MRC Biostatistics Unit, CAMBRIDGE CB2 0SR
Bird, SM, MRC Biostatistics Unit, CAMBRIDGE CB2 0SR
Merrall, ELC, Novartis Vaccines and Diagnostics, Amsterdam
Hutchinson, SJ, Health Protection Scotland, Glasgow G3 7LN

For the follow-up era of 2006-2010, we validated the earlier finding by Merrall et al. (2012) that the 28-days after hospital-discharge are a period of high risk for drugs-related death (DRD) among drug treatment clients in Scotland. Secondly, using the combined data for 1996-2010, we show that the behavioural risk-factor of having ever injected is far superior to length-of-stay or discharge-diagnosis as a basis for identifying those at highest DRD-risk in the 28-days after hospital discharge.

Methods: Linkage of hospitalization and death records to the Scottish Drugs Misuse Database (SDMD) resulted in a 'virtual' cohort of over 98,000 individuals who registered for drug treatment in Scotland during 1 April 1996 to 31 March 2010, and contributed 705,538 person-years (pys) of follow-up, 173,107 hospital-stays, and 2,523 DRDs. Time-at-risk of DRD was categorized as: during hospitalization, within 28 days, 29-90 days, 91 days -1 year, >1 year since most recent hospital discharge versus 'never admitted' (reference category). Factors of interest were: length of hospital-stay (0-1 versus 2+ days), main discharge-diagnosis and having ever injected.

Findings: First, we confirmed SDMD clients' high DRD-rate (per 1,000 pys) soon after hospital-discharge in 2006-2010, namely: 50 (95% CI: 39-63) during hospitalization; 23 (19-27) within 28 days, 12 (9.9-14.5) during 29-90 days and 8.6 (7.5-9.8) during 91 days to 1 year after discharge versus 3.9 (3.5-4.3) when > 1 year after most recent hospitalization and 1.2 (1.1-1.4) for those never admitted.

Secondly, during 1996-2010, DRD-rate in the 28 days after hospital-discharge did not vary by length of hospital-stay but was very significantly higher for SDMD clients who had ever-injected: 32 (95% CI: 27.7-36.2) versus 13 (10.0-16.4) for those who were not known to have injected. High DRD-rates within 28 days were observed for main discharge-diagnoses of poisoning by drugs, medicaments and biological substances, 44 (34-57), and mental and behavioural disorders excluding psychoactive substance misuse, 36 (27-47). However, these two diagnoses accounted for only 113/290 (39%) DRDs observed in the 28 days after hospital-discharge whereas ever-injectors accounted for 222/290 (77%).

Interpretation: Hospital-discharge marked a period of increased DRD vulnerability in 2006-2010 as in 1996-2006, regardless of length of hospital-stay, and especially for those with a history of injecting.

Funding: Medical Research Council, Chief Scientist Office of the Scottish Executive

Corresponding author email: sheila.bird@mrc-bsu.cam.ac.uk

Delivering a UK Research Platform

thompson, s, Swansea University, SAIL

Research funders require grant-holders to write data management plans that incorporate data preservation and sharing. Yet, little progress has been made. It is still very difficult to access linked data from existing cohorts. There are many underlying issues, including ownership, threats of deductive disclosure even when data are de-identified, the sheer volume and complexity of datasets and the lack of facilities and health informatics expertise in those managing cohorts or wishing to analyse linked data. Our experience from working on data linkage with ALSPAC, UK Biobank and Millennium Cohort and on the design of LifeStudy has provided an in-depth insight into these challenges and their potential solutions. The data are simply too complex and large scale to be moved easily and without considerable information and research governance risks.

We have developed a secure e-research platform (SeRP) to enable safe access to large numbers of researchers to the complex multi-dimensional datasets that comprise modern cohorts, based around an expansion of the technology successfully used in pilots with the SAIL Gateway.

SeRP will be offered as a national resource. It will devolve account control via management controls on the system to any bona fide project, giving full information governance responsibilities and control to the projects / data owners and not to the platform or to the Institute. Hence, custodianship and access control will remain with the original data owners. SeRP is not designed to just be a datastore or linkage facility. Rather, it will provide functionality to the owners of both to provide access to subsets of the data and supportive tools (metadata libraries, knowledge bases, communication and collaboration tools). Adopting an open standards approach will provide for growth in analytical functionality, powered by high performance computing and produce an attractive, efficient, cost-effective, and powerful research environment to access such data safely.

Corresponding author email: simon@chi.swan.ac.uk

Secure Unified Research Environment: watch it work!

Khoo, J, Sax Institute, Sydney AUSTRALIA
Churches, T, Sax Institute, Sydney AUSTRALIA

The Secure Unified Research Environment (SURE) is a remote-access computing facility that has been built to provide secure access to linked health-related data for research. SURE is operated by the Sax Institute, and is funded by the Australian and New South Wales Governments as part of the Population Health Research Network. SURE was officially launched in July 2012 and as at April 2013, had more than 30 active users.

Although developed to meet the needs of Australian researchers within the Australian policy and funding context, SURE shares features with "safe haven" and "data enclave" initiatives in other countries that have been driven by similar imperatives. It also has unique features, including:

- purpose-built software that manages movement of files and data into and out of the facility, while allowing for project-specific policies regarding who can approve this;
- researcher workspaces that are highly customisable in terms of the software that can be deployed and other technical specifications;
- workspaces are partitioned in such a way that data from two project workspaces cannot be combined, even if a researcher has access to more than one project workspace; and
- tools to enhance collaboration and productivity of researchers can be deployed within workspaces such as version control programs and wikis.

This presentation will involve a "live" demonstration of SURE, including examples of the sophisticated methods of analysis that are able to be implemented in the facility as a result of its design. Following the demonstration, a facilitated panel session involving members of the SURE team, and researchers who are currently using the facility, will allow participants to explore in detail the features of SURE and the experiences of its users.

Corresponding author email: joanna.khoo@saxinstitute.org.au

Development of National Linkage Infrastructure in Australia

Boyd, JH, Curtin University

The Centre for Data Linkage (CDL) at Curtin University in Pert, Western Australia was established to enable linkage of state and national datasets to service researchers. In order to do so effectively, the CDL has developed a series of compatible tools and modules in a linkage infrastructure framework.

Within this infrastructure framework, the CDL have developed an end-to-end production system to manage linkages and extractions, which handles the more complex processes typically found in a large scale production linkage unit. The CDL have also implemented a secure file transfer system to enable data delivery to and from this service. Finally, CDL have identified, and implemented a secure infrastructure system, designed to host this software, while ensuring security, privacy and the meeting of all legislative requirements.

This talk will focus on the process of designing and implementing this infrastructure, along with some of the key features found in each system.

Corresponding author email: J.boyd@curtin.edu.au

An approach to multi-jurisdictional wide-scale data linkage across Australia

Boyle, DIR, University of Melbourne Rural Health Academic Centre

Background

Whilst most data linkage activities in Australia involve linkages between large organisational datasets with foundations rooted in MOU's or government legislation, such a model is difficult to expand to linkage encompassing hundreds or thousands of GP and community data resources. In an aim to support linkage in Australia across all jurisdictional boundaries regardless of size or location the University of Melbourne has developed the GRHANITE software system.

Methods

The following issues needed to be addressed and solutions have been designed into the solution:

- A generic solution cannot rely on data standards - there are few widely implemented standards in Australia
- The system must interface to data extracts or interrogate directly (small organisations do not have the knowledge to interrogate their databases)
- Study protocols need enforced to ensure only permitted data leaves custodian organisations
- In many cases patient identifiers cannot be released - linkage information must be privacy-protected before data export
- Consent (opt-in, opt-out, waiver) must be supported controlling data release
- The system must support hundreds of locations requiring automation in data extracts, software installation, updates and study protocol updates
- The system must support multiple projects delivering data to multiple organisations (install once - use many)
- Performance and security must be such that any scale of project can be undertaken and guarantees given

GRHANITE meets these design briefs and implements itself through a configurable installer and an XML and SQL based language that defines database connectivity, study protocols and data linkage parameters. The system includes an Australian dataset verified privacy-protecting record linkage system. Central web services coordinate all management activities: software updates, protocol releases and updates and data distribution.

Results

GRHANITE is installed in over 200 GP, Laboratory, Hospital and other community sites routinely extracting data weekly. Over 500 new installations are planned and 8 major audit and research studies utilise the system. The privacy-protecting linkage system is overcoming legislative restrictions and permission to link with large resources such as National Registers and National Death Data has been granted.

Discussion and Conclusions

With a fragmented and largely private health system and almost no standards accreditation for system suppliers, systematically tackling data linkage in Australia is challenging. Whilst hundreds not thousands of sites have been linked, the GRHANITE tool is already Australia's most prevalent tool.

Corresponding author email: dboyle@unimelb.edu.au

Compare hospital variability in acute myocardial infarction care and outcome between Sweden and the UK

Chung, SC, Research Department of Epidemiology & Public Health, UCL
Sundstr.m, J, The George Institute and Uppsala University Hospital
Nicholas, O, National Institute for Cardiovascular Outcomes Research, UCL
James, S, Uppsala Clinical Research Center, Uppsala University
Jeppsson, A, Department of Molecular and Clinical Medicine, Institute of Medicine, Sahlgrenska Academy, University of Gothenburg
Deanfield, J, Centre for Cardiovascular Prevention and Outcomes, UCL
Timmis, A, National Institute for Health Research, Biomedical Research Unit, Barts Health London
Jernberg, T, Department of Medicine, Huddinge, Section of Cardiology, Karolinska Institute
Hemingway, H, Research Department of Epidemiology & Public Health, UCL

Background: Between-hospital variability in mortality after acute myocardial infarction (AMI) provides a useful insight into the effectiveness of national healthcare systems. Better performing systems should deliver both improved mean mortality and less variability between hospitals. Between-country comparisons of hospital variability have not previously been reported for AMI.

Objectives: We sought to compare Sweden and the UK in relation to the (i) hospital variability in AMI treatment and 30-day mortality, (ii) time trends in hospital variability, and (iii) the association between hospital variability in 30-day mortality and hospital treatment use at study entry, and change in hospital treatment from the previous year.

Methods: We used nationwide registry data with 100% coverage for hospitals providing AMI care for in Sweden (Register of Information and Knowledge about Swedish Heart Intensive care Admissions, SWEDEHEART/RIKS-HIA) and the UK (Myocardial Ischaemia National Audit Project, MINAP) between 2004 and 2010 (n=87 in Sweden and n=242 in the UK). Hospital-specific proportions on case-mix, treatment measures were generated based on patient data (n=119,786 in Sweden and n=391,077 in the UK). The primary outcome was death from any cause at 30-day after AMI admission, measured at patient level and analysed at hospital level. All data were analysed in London according to common protocol via secure remote access.

Results: The mean and standard deviation of 30-day mortality across all hospitals was 7.6% (2.8%) in Sweden and 10.3% (3.9%) in the UK. The mean hospital 30-day mortality decreased from 2004 to 2010 from 10.1% to 6.2% in Sweden, and 12.2% to 8.5% in the UK. The hospital use of primary PCI increased in both countries from 2004 to 2010, from 26% to 62% in Sweden and from 4% to 20% in the UK. Over the same period the variability among hospitals in the use of primary PCI also increased in both Sweden (from 24% to 28%) and the UK (from 14% to 33%). In Sweden, the between hospital variability in discharge medication was lower than in the UK. In multivariate models, higher rates of hospital reperfusion therapy use in 2004 and improvement from the previous year were associated with lower 30-day mortality. The associations were independent of casemix.

Discussion: There was higher variability between hospitals in AMI treatment and 30-day mortality in the UK as compared to Sweden. Understanding the reasons for these system-wide differences may help improve the quality of care and outcomes of acute myocardial infarction.

Corresponding author email: s.chung@ucl.ac.uk

Reporting of congenital cardiac interventional procedures: anonymous record linkage between a mandatory register and an administrative data set for capture-recapture analysis

Burn, J, Newcastle upon Tyne Hospitals NHS Foundation Trust
Keltie, K, Newcastle upon Tyne Hospitals NHS Foundation Trust
Bousfield, DR, Newcastle upon Tyne Hospitals NHS Foundation Trust
Colechin, ES, Newcastle upon Tyne Hospitals NHS Foundation Trust
Patrick, H, National Institute for Health and Care Excellence
Cunningham, D, National Institute for Cardiac Outcomes Research, UCL
Sims, AJ, Newcastle upon Tyne Hospitals NHS Foundation Trust

Objective

To estimate the completeness of the Central Cardiac Audit Database (CCAD) for four cardiac interventional procedures using a capture-recapture approach, by matching mandatory entries in CCAD with a routine administrative dataset for NHS episodes in England, Hospital Episodes Statistics (HES).

Method

Four procedures were studied, for which NICE guidelines were published and data submission to CCAD was mandatory (balloon dilatation of pulmonary valve stenosis, IPG67; endovascular closure of atrial septal defect, IPG96; endovascular closure of patent ductus arteriosus, IPG97; percutaneous pulmonary valve implantation for right ventricular outflow tract dysfunction, IPG237). Episodes of care between April 2006 and March 2010 for each procedure were identified from HES using OPCS-4 codes recommended by the NHS Classification Service. Procedure records were extracted from CCAD using European Paediatric Cardiac Codes (EPCC) and other filters which most closely matched the procedure definitions. An anonymous record-linkage algorithm was developed to identify cases common to both anonymous data sets. Linkage was based on provider (hospital), year and month of main procedure, age at procedure (whole years) and gender; episodes with missing values from any of the five matching fields were excluded. For each procedure, the number of cases common to both data sets, the number in HES only and the number in CCAD only were estimated. The total number of each procedure type (in England, April 2006 to March 2010) and completeness of each data set were estimated using a capture-recapture approach, assuming independence of reporting to the two data sets.

Results

For the four procedures (IPG67, IPG96, IPG97, IPG237) the proportions of episodes in HES for which matches in CCAD were found were 86.6%, 67.8%, 89.9% and 38.2%. The proportions of records in CCAD for which matches in HES were found were 61.9%, 48.5%, 78.3% and 33.7%. By capture-recapture analysis, the estimated total number of procedures were 996, 3294, 1974 and 238, with completeness estimated as 84.7%, 67.8%, 90.0%, 38.7% (CCAD) and 60.5%, 48.5%, 78.3% and 34.0% (HES). The proportions of records with missing data in the fields used for matching was 4.7%, 1.8%, 1.3%, 22.7% (CCAD) and 15.4%, 5.9%, 6.0% and 31.4% (HES).

Discussion

Linkage of anonymous data sets and application of capture-recapture analysis was useful in identifying procedures with poor completeness in CCAD or mis-coding in HES (e.g. IPG237 in this study). Further searches of HES will be combined with anonymous linkage and capture-recapture analysis to separate coding errors from under-reporting.

Corresponding author email: andrew.sims@nuth.nhs.uk

National cardiovascular registries for measuring hospital variation in mortality after myocardial infarction: Cardiovascular disease research using Linked Bespoke studies and Electronic health Records (CALIBER)

Timmis, A, NIHR Biomedical Research Unit, Barts Health NHS Trust, London
Nicholas, O, Epidemiology & Public Health, UCL
George, J, Epidemiology & Public Health, UCL
Eldridge, S, Centre for Statistics, Queen Mary University, London
Feder, G, School of Social and Community Medicine, University of Bristol, Bristol
Hemingway, H, Epidemiology & Public Health, UCL

National cardiovascular registries for measuring hospital variation in mortality after myocardial infarction: Cardiovascular disease research using Linked Bespoke studies and Electronic health Records (CALIBER)
Background: Hospital variation in patient outcomes might provide insight into overall system performance, based on the null hypothesis that there should be no substantial variation in outcomes within an equitable health service. We have tested this hypothesis using the national registry of patients admitted with acute myocardial infarction (the Myocardial Ischaemia National Audit Project (MINAP) registry) which records data on case-mix, treatment and mortality

Objectives: To determine whether there is clinically important variation in 30-day mortality rates for acute myocardial infarction after adjustment for case mix and treatment provided in individual hospitals and, if variation exists, whether underperforming hospitals can be reliably identified.

Methods: Prospective cohort study in all 236 hospitals in England and Wales of 316,648 consecutive patients with first ST elevation myocardial infarction or non-ST elevation myocardial infarction recorded in MINAP between 2003 and 2009. The primary outcome was the 30-day all-cause mortality rate for individual hospitals. We ranked hospitals by the fully adjusted 30-day mortality and ascertained the fraction of hospitals conferring excess risk (hazard ⩾40% above the mean for all hospitals, equivalent to increasing patient age by at least five years) using random effects models with multiple imputation of missing data.

Results: Crude 30-day mortality for the complete study population was 9.4%. Following adjustment for case mix, acute and discharge treatment and hospital level factors, the mean standard deviation of log-hazard across all hospitals was 0.24 (0.23-0.25), significantly exceeding the assumption of no variation (p<0.001). Hospitals with a 30-day mortality >1 standard deviation either side of the mean added or subtracted at least 21% (1-exp (-0.24)) to the average risk. Fifteen hospitals (6.4% (95% CI 4.7% to 8.5%)) met the ⩾40% criterion for excessive 30-day mortality, but in ranking all hospitals by fully adjusted 30-day mortality, the 95% confidence intervals around the estimates were wide, and in only 2 hospitals was there a >90% probability of being in the excess mortality category. The variability in hospital 30-day mortality was comparable with the overall improvement in 30-day mortality over the study period. We are now extending this research internationally and comparing the extent of hospital variation between UK and Sweden (the only other country with a national ACS registry) in order to provide evidence for quality of care improvement initiatives.

Conclusions: About one in 20 hospitals treating patients with acute myocardial infarction in England and Wales have a 30-day mortality risk equivalent to increasing patient age by at least five years. Underperforming hospitals cannot be identified with precision but the variation between hospitals is serious and only partially explained by variation in use of effective treatment.

Corresponding author email: adamtimmis@mac.com

National Registries for Evaluating Acute Coronary Syndrome Outcomes: MINAP experience from the National Institute for Cardiovascular Outcomes Research (NICOR)

Gale, CP, Centre for Epidemiology and Biostatistics, Leeds
Simms, AD, Centre for Epidemiology and Biostatistics, Leeds
West, RM, Centre for Epidemiology and Biostatistics, Leeds
Fox, KAA, Centre for Cardiovascular Science, University of Edinburgh
Timmis, A, NIHR Biomedical Research Unit, Barts Health NHS Trust, London
Hemingway, H, Epidemiology & Public Health, UCL
Deanfield, JE, Centre for Cardiovascular Prevention and Outcomes, UCL

Background
Inequalities in hospital acute coronary syndrome (ACS) outcomes necessitate improved indicators of quality of care - to evaluate the full ACS treatment pathway and to quantify variation in outcomes after standardisation for case-mix.

Methods
The Myocardial Ischaemia National Audit Project (MINAP) is a prospective whole country registry of around 1 million patients hospitalised with ACS from 2000 through 2013. Each case entry offers rich details of the patient journey. All-cause mortality is tracked through linkage to the National Health Service Central Register (NHSCR) using a unique National Health Service (NHS) number. Established ACS risk scores were evaluated for use with MINAP data, missed opportunities for care quantified and composite scores of performance developed. The impact of missing data and different statistical approaches to outcomes reporting were investigated.

Results
Mortality rates after ACS were low and indicators of hospital quality of care were high. There was, however, considerable variation in outcomes not attributable to key patient-level factors, supporting the notion of regional-dependent variation in ACS care - increasing hospital volume of treated cases was associated with reduced mortality. Furthermore, there was evidence for missed opportunities to provide ACS care. These shortfalls in care were more frequent in females, the elderly and patients at higher risk of mortality and were associated with subsequent missed opportunities for care, and early and late mortality. The Global Registry of Acute Coronary Events (GRACE) risk score demonstrated good model performance across a range of indices using MINAP data - allowing standardisation for case-mix. A multidimensional hospital composite score discriminated hospital performance more readily than single indicators, and was significantly inversely associated with mortality.

Discussion
The MINAP registry of patients hospitalised with ACS in England and Wales provides a unique and internationally competitive opportunity to evaluate ACS outcomes in the NHS. Mortality after ACS is low, but regional variation in care is evident. Modifiable factors associated with mortality and target populations in whom care may be optimised are readily identifiable. The use of routine electronic health for health service research will allow healthcare professionals and regulators to improve outcomes for patients with ACS.

Corresponding author email: c.p.gale@leeds.ac.uk

Linking of primary care records to census data to study the association between socio-economic status and health outcomes: a nation-wide ecological study.

Garcia-Gil, MM, System for the Development of Research in Primary Care. Institut d'Investigació en Atenció Primària. IDIAP Jordi Gol. Catalonia, Spain.
Elorza, JM, System for the Development of Research in Primary Care Institut d'Investigació en Atenció Primària. IDIAP Jordi Gol. Catalonia, Spain.
Alves-i-Cabratosa, L, Vascular Health Research Group. Institut d'Investigació en Atenció Primària. IDIAP Jordi Gol. Catalonia, Spain.
Hermosilla, E, System for the Development of Research in Primary Care. Institut d'Investigació en Atenció Primària. IDIAP Jordi Gol. Catalonia, Spain.
Et al.

Introduction: Construction and validation of census deprivation indices overcomes the absence of socioeconomic data in primary care electronic medical records. The Spanish MEDEA index is a composite of 5 variables (unemployment, manual work, temporary work, insufficient education, and insufficient education among young people) obtained from census information.

Objective: To study the association between the Spanish MEDEA deprivation index and relevant co-morbidities in the SIDIAP database (www.sidiap.org).

Methods SIDIAP contains anonymised longitudinal patient information including demographic characteristics, morbidity (ICD-10), lifestyle variables, laboratory tests, and pharmacy invoice data) for >5 million people (>80%) from Catalonia (Spain). Main exposure was the ecological MEDEA index, calculated using census-based socioeconomic indicators. SIDIAP was linked to the 2001 census after harmonization of the census residences/addresses information. Rural areas were excluded, as MEDEA has only been validated in urban areas. The study outcomes were incident cases of: acute myocardial infarction (AMI), stroke, osteoarthritis (hand, hip and knee), rheumatoid arthritis, and cancer (cervix, uterus, breast, colon, prostate, and lung). The association between MEDEA quintiles (the higher, the more deprived) and the incidence of these outcomes in 2009-2012 was evaluated using age and gender-adjusted zero-inflated Poisson regression. We further adjusted for lifestyle factors (smoking, alcoholism, obesity), and long-term common co-morbidities (hypertension, diabetes) to study potential causal pathways.

Results: Sex-age adjusted models showed a significant association between MEDEA quintiles and incident AMI (RR 1.13 [95%CI 1.06-1.20] for most deprived compared to most affluent), stroke (RR 1.24 [1.19-1.30]), osteoarthritis (RR 1.57 [1.53-1.62] for knee, 1.31 [1.20-1.44] for hand, and 1.24 [1.18-1.29] for hip osteoarthritis); lung, cervix and uterus cancer (RR 1.30 [1.20-1.40], RR 1.22 [1.08-1.39] and RR 1.24 [1.12-1.36], respectively). Conversely, an inverse association was seen with breast, prostate and colon cancer (RR 0.79 [95%CI 0.75-0.84], RR 0.74 [95%CI 0.69-0.79] and RR 0.90 [95%CI 0.84-0.95]) and rheumatoid arthritis RR 0.88 [95%CI 0.79-0.97]. All these associations (but AMI, and osteoarthritis at hand and hip) stood for multivariate adjustment.

Conclusions: Socio-economic status is associated in different directions with the incidence of a number of relevant conditions in SIDIAP. The observed associations (eg highest incidence of breast cancer in the most affluent, versus highest risk of lung cancer in the most deprived) are supportive with existing literature. The linkage of SIDIAP and census data has proved essential to ascertain the excess risk of health outcomes related with social deprivation in primary care.

Corresponding author email: lalves@idiapjgol.info

Socio-economic position and childhood multimorbidity: a study using linkage between the Avon Longitudinal Study of Parents and Children and the General Practice Research Database

Cornish, RP, School of Social and Community Medicine, University of Bristol
Boyd, AW, School of Social and Community Medicine, University of Bristol
Salisbury, C, School of Social and Community Medicine, University of Bristol
Van Staa, T, CPRD, MHRA
Macleod, J, School of Social and Community Medicine, University of Bristol

Background

In adults, multimorbidity is associated with social position. Socially disadvantaged adults typically experience more chronic illness at a younger age than comparable individuals who are more advantaged. The relation between social position and multimorbidity amongst children and adolescents has not been as widely studied and is less clear.

Methods

The NHS Information Centre linked participants in the Avon Longitudinal Study of Parents and Children to the General Practice Research Database. Multimorbidity was measured in three different ways: using a count of the number of drugs prescribed, a count of chronic diseases, and a person's predicted resource use score; the latter two measures were derived using the Johns Hopkins ACG system. A number of different socio-economic position variables measured as part of ALSPAC during pregnancy and early childhood were considered. Ordered logistic and negative binomial regression models were used to investigate associations between socio-economic variables and multimorbidity.

Results

After mutually adjusting for the different markers of socio-economic position, there was evidence, albeit weak, that chronic condition counts among children aged from 0 to 9 years were higher among those whose mothers were less well educated (OR=0.44; 95% confidence interval 0.18-1.10; p=0.08). Conversely, children whose mothers were better educated had higher rates of chronic illness between 10 and 18 years (OR = 1.94 ; 95% CI 1.14-3.30). However, living in a more deprived area, as indicated by the Townsend score, was associated with a higher odds of chronic illness between 10 and 18 years (OR for each increasing decile of Townsend score = 1.09; 95% CI 1.00-1.19; p=0.06).

Conclusions

We have found that, in younger children, multimorbidity is higher amongst children whose parents are less well educated. In older children and adolescents this association is less clear and, according to some measures, multimorbidity appears higher among those who are more advantaged. We have also demonstrated that linkage between prospective observational studies and electronic patient records can provide an effective way of obtaining objectively measured outcome variables. The sample size in this study was quite small and we would therefore recommend further research is done to replicate these findings. With a larger dataset it would also be possible to try to determine the pathways through which social position might be impacting on multimorbidity.

Corresponding author email: rosie.cornish@bristol.ac.uk

Effect of socioeconomic and health inequalities on mortality: The Turin Longitudinal Study

Rasulo, D, Regional Epidemiology Unit, ASL TO3 Piedmont Region, Turin, Italy
Costa, G, Department of Clinical and Biological Sciences, University of Turin, Italy
Spadea, T, Regional Epidemiology Unit, ASL TO3 Piedmont Region, Turin, Italy

It is well known that there is a socioeconomic gradient in mortality, namely individuals higher in the social hierarchy enjoy better health than do those below. The effect of socioeconomic inequalities can be direct, in the sense that are attributable to personal attributes, or can be tied to the resources of the family of origin. At the same time, the health of adults appears to be influenced by parents' longevity. Hence, taken as a whole, the literature points to the possibility that socioeconomic resources combine with parental socioeconomic and health inequalities to influence mortality.

Our study contributes to an understanding of the role of socioeconomic and health inequalities on mortality using the Turin Longitudinal Study (TLS) which facilitated a child-parent matched analysis. The Turin Longitudinal Study is a record-linkage study containing information from censuses and routinely registered events in Turin such as date and cause of death. It began with the 1971 census when it enrolled all the Turin residents who were present at the census date. The algorithm which created the child-parent matched dataset used the Turin Population Register and the census household form.

The study verified whether, over a three-decade period, the effect of socioeconomic conditions was mediated by the conditions of the family of origin and whether parental life span was an independent factor for mortality. We estimated Poisson hazard models to evaluate the ways in which socioeconomic and health inequalities were associated with mortality. To model socioeconomic differences in mortality, we considered time-varying measures of education, employment, housing typology, marital status and parliamentary constituencies' deprivation. The same variables provided an array of conditions for the family of origin. The role of health inequalities in the family of origin was assessed through the effect of parental life span on mortality.

Drawing on the Turin Longitudinal Study, the analysis showed that mortality for both sexes was associated with all socioeconomic variables attributable to personal attributes as well as to the highest educational attainment in the family of origin. In particular, a stratified variable indicated that the effect of education on mortality was mediated by parental education. On the other hand, health inequalities in the family of origin were independently associated with mortality as indicated by a decreased risk with increasing parental life span. Our study then indicates that, though mortality is largely associated with adult social conditions, social and health disadvantage in the family of origin plays a significant role on mortality.

Corresponding author email: domenica.rasulo@ons.gsi.gov.uk

Linking Healthcare Associated Infection Data to Patient Episode Data to Measure the Cost of HAI to NHSScotland

Cairns, S, National Services Scotland Health Protection Scotland
Bishop, J, National Services Scotland Information Services Division
Fleming, S, National Services Scotland Health Protection Scotland
Reilly, J, National Services Scotland Health Protection Scotland
Robertson, C, University of Strathclyde, Health Protection Scotland
Walker, A, University of Strathclyde

Healthcare associated infections (HAI) are a significant threat to patient safety worldwide and contribute to avoidable harm in hospital patients. HAI place a significant economic burden on healthcare systems. Understanding the costs associated with healthcare associated infection is essential to provide justification for continued investment in infection prevention and control (IPC) measures and in ensuring the cost effectiveness of IPC programmes.

The aim of this study was to estimate the cost of HAI originating in acute hospital care inclusive of costs associated with extended lengths of stay in hospital following discharge from the hospital where the HAI originated.

A rolling point prevalence survey of healthcare associated infection and antimicrobial use was carried out in Scottish hospitals in September and October 2011. The PPS dataset was linked to the Scottish Morbidity Record (SMR-01 and SMR-04). The SMR records diagnosis data including date of discharge when a patient is discharged from hospital, changes consultant or is transferred to another hospital enabling the total length of stay to be determined. Linkage to SMR enabled the date of discharge from the survey hospital and date of discharge from any subsequent transfer hospitals to be determined. This enabled the total length of stay from date of admission to the survey hospital to final discharge from hospital to be determined.

The length of stay data were analysed using parametric survival analyses. Several models were created to model the length of stay. Competing models were assessed using the Akaike Information Criterion. The final model included HAI status, specialty of patient care and McCabe Score. McCabe Score measures the underlying medical condition of the patient and inclusion in the model prevents confounding by the patients' underlying condition. The predicted lengths of stay by specialty were calculated as a weighted average of the individual estimated median stays by HAI status, specialty and McCabe score using the proportion of patients in each McCabe category as the weight.

The additional length of stay in patients with HAI was used to calculate the cost of HAI using costing data from the Scottish Health Service Costs. The additional length of stay due to HAI and the cost of HAI to NHSScotland overall and by specialty of patient care will be presented (results TBC).

Corresponding author email: shona.cairns@nhs.net

The Kuwait Health Network: utilising routinely collected data to support quality improvement of diabetes care in Kuwait

Conway, N.T, University of Dundee
Al Khuzam, A.F, Dasman Diabetes Institute, Kuwait
Al Ali, S., Kuwait Ministry of Health
Al Faraj, J., Kuwait Ministry of Health
Al Mansour, S., Kuwait Ministry of Health
Al Musallam, S., Kuwait Ministry of Health
Bell, A., Aridhia Informatics Ltd.
Judson, A, Aridhia Informatics Ltd.
Kelly, C, Aridhia Informatics Ltd.
Sibbald, D, Aridhia Informatics Ltd.
Wake, D.J, University of Dundee
Behbehani, D.J, University of Dundee

Introduction: The Kuwait Health Network (KHN) is the product of an international collaboration between NHS Tayside, University of Dundee, Kuwait Ministry of Health, Dasman Diabetes Institute, and Aridhia Informatics Ltd to support the management of patients with diabetes living in Kuwait - approximately 24% of the population. KHN is an outcome-focused analytical application that provides healthcare professionals (HCP's) with instantly accessible information regarding current provisions of care. We aim to illustrate how the analytical capabilities of KHN can utilise routinely collected healthcare data to provide an epidemiological snapshot of those living with diabetes in Kuwait.

Methods: KHN v1.0 was implemented in February 2013 in four primary health centres within the Kuwait capital region. A secure network with both role and location-based access links laboratory data with a disease registry based on a defined dataset of coded clinical terms. KHN reports on Quality Performance Indicators (QPI's) based on Kuwait national clinical standards for diabetes. These include the recommendation that HbA1c is measured at least 6 monthly; cholesterol and body mass index (BMI) are measured annually; and blood pressure (BP) measured at each clinic visit. A fully automated report generator utilising a combination of analytical and document-preparation freeware can instantly collate domain-specific results, thereby providing the ability to easily publish and disseminate a near real-time summary of QPIs in a print/tablet-friendly format.

Results: By April 2013, there were 4390 registered patients, 4305 (98.1%) of whom have type 2 diabetes. Within the last 15 months: 3006 (68.4%) had HbA1c measured; 1645 (37.5%) had cholesterol measured; 2026 (46.1%) had BMI recorded; and 2721 (62%) had BP recorded. Of these patients: 25.8% (95%CI:24.3-27.4) had HbA1c within the target range (<53mmol/mol); 81.2% (95%CI:79.2-83.1) had cholesterol within the target range (<5.2mmol/l); 25.7% (95%CI:23.8-27.7) had BMI within the healthy range (18-25kg/m2); and 37.0% (95%CI:35.2-38.8) had a normotensive BP (<130/80mmHg). Planned KHN software releases for the coming months include: additional QPI's; disease stratification; and clinical decision support tools.

Conclusion: These results suggest that large numbers of patients with diabetes in Kuwait are not meeting national clinical standards. If representative, this would suggest that the Kuwaiti diabetic population is at significant risk of developing future disease complications. Optimising care requires HCP's to identify and predict at-risk patients and share this information effectively across all healthcare sectors. By meeting this need, KHN also provides robust population level data which is instantly accessible for quality improvement, healthcare planning or research purposes.

Corresponding author email: n.z.conway@dundee.ac.uk

Primary care research using national, regional and organization database record linkage in New Zealand: A foundation for international research?

Dovey, SM, University of Otago, New Zealand
Tomlin, A, BPACNZ
Lloyd, H, BPACNZ
Loh, LW, University of Otago and BPACNZ
Tilyard, MW, University of Otago and BPACNZ

New Zealand has national collections of routinely collected health data on all hospital admissions, medicines dispensed in the community, investigations conducted in community laboratories, general practice consultations, births, deaths, and some specific disease registries (e.g. cancer and diabetes). Networks of general practices also contribute their full patient records to research databases. All databases are linkable using a unique patient identifier, the National Health Index (NHI) code. The NHI also provides demographic and socioeconomic data. The aim of this presentation is to explore opportunities and challenges for international collaborative health research. We use two examples of recent research using health databases in New Zealand and borrowing from or collaborating with researchers using similar data from other countries.

First, we present findings from our replication of the Scottish (University of Dundee) Predicting Emergency Admissions Over the Next Year (PEONY) model. This modelling project enabled risk stratification of all general practice patients in New Zealand using data from three national datasets: the National Minimum Dataset for Hospital Events, the Primary Health Organization enrolment registers, and the Pharmaceutical Collection of dispensed medicines. We retained all variables used in the PEONY model and calculated new regression coefficients for New Zealand patients >40 years. The model with best predictive power for New Zealand limits the number of admissions and total bed days and the number of items dispensed to maximum value of 100 over three years. We added patient ethnicity to the model and this marginally improved its performance.

Second, we explored the alignment of routine New Zealand health data for patients with diabetes with similar data from Ontario, Canada to compare diabetes prevalence. We identified 155,472 patients from dispensed medicines and a further 4,182 patients hospitalised with diabetes. The total number of patients identified from these two sources was 159,654 (3.8% of New Zealand population) but this estimate includes only patients treated with medication or hospitalised. The Ontario Diabetes Database cohorts of diabetes patients are also identified from mutiple sources. They also use hospital discharge data but have the crucial difference of having a physician billing database containing diagnosis codes for diabetes. The entire Ontario Diabetes Database is recreated yearly using updated physician claims and hospital discharge data and incorporates Registered Persons database information. Age and sex adjusted diabetes prevalence in Ontario is estimated at 10.6%.

Increasing familiarity of researchers with routine health databases provides new opportunities for collaborative research that may provide innovative insights into the performances of health systems internationally.

Corresponding author email: susan.dovey@otago.ac.nz

How do I love data? Let me count the ways....Using existing databases for work and health research

Koehoorn, M, School of Population and Public Health, University of British Columbia
McLeod, CB, University of British Columbia

In this electronic age, many large databases have been collected for purposes other than research, such as administrative health billing records, cancer registries and insurance claims. We developed an innovative partnership between policy-makers in the workers' compensation system and university researchers in occupational and environmental health to maximize the use of these data for research purposes (Partnership for Work, Health and Safety). Using examples of projects in the areas of surveillance, epidemiology and policy evaluation, this presentation will highlight the research potential (and challenges) of using existing databases for policy relevant research questions in the area of work and health.

For surveillance for example, workers' compensation claims were linked with cancer registry records for individuals diagnosed with mesothelioma in the Canadian province of British Columbia, demonstrating that individuals with an occupational cancer were not seeking compensation benefits. This finding has led to several interventions to increase awareness of mesothelioma as a compensable occupational cancer (e.g. the use of the cancer registry to send a letter to physicians of patients with mesothelioma) and awareness of workers' compensation procedures (e.g. development of mobile apps to facilitate submission of physician reports to the compensation system). For epidemiology for example, workers' compensation claim data was linked with medical services health records and a job exposure matrix for workplace allergens, demonstrating that work-aggravated asthma was under-compensated in British Columbia. This finding, as part of a body of evidence, led to a change in policy recognizing work-aggravated asthma as a compensable condition. For policy evaluation for example, workers' compensation claim data was linked with medical services data for workers' undergoing knee surgery in private clinics versus public hospitals, demonstrating no significant difference in return-to-work outcomes by surgical setting. As another example of policy evaluation, workers' compensation claim data was linked with data from a provincial safety association on forestry fallers (a high risk occupation), demonstrating that the occupational certification program has not significantly reduced the risk of workplace injuries. Both of these policy evaluations led to refinements of system procedures to improve the delivery of prevention and rehabilitation services.

The Partnership's research is designed to optimize development and linkage of workers' compensation claim data for research purposes. As a result of the active engagement of the stakeholders in the research project, the data has been used for policy relevant research questions and the findings have informed interventions, policy changes and program refinements for the health, safety and rehabilitation of workers. The success of using administrative linked data by the Partnership have led to newly proposed projects investigating comparable compensation indicators across national and international jurisdictions, gender differences in return-to-work outcomes, and patterns of opioid usage among workers following implementation of clinical guidelines.

Corresponding author email: mieke.koehoorn@ubc.ca

Transforming a natural experiment into a health intervention evaluation using linked routine data

Rodgers, S.E, College of Medicine Swansea University
Heaven, M, College of Medicine Swansea University
Lacey, A, College of Medicine Swansea University
Dunstan, F, School of Medicine Cardiff University
Demmler, J, College of Medicine Swansea Univerisity
Poortinga, W, School Of Physchology Cardiff University
Lyons, R, College of Medicine Swansea University

A cohort comprising residents of a housing regeneration and health programme was created from routinely collected data using a system which allows us to anonymously link housing data to individuals and their health (Rodgers et al. 2012). The regeneration programme, consisting of four rolling work packages, runs from 2009 to 2014. The baseline cohort comprises 16,336 residents of the 9051 social housing residences who had a full year of residence history in 2009. We have refined our method to include a count of “days at risk”, capturing days in a study residence. This allowed us to include individuals with less than a year of residency in the study, and increased the number of residents in the cohort by 21.6% to 19,860 (1st Jan 2009 to 31st December 2009). We recorded the duration of residence for individuals who spent time outside the study social houses and calculated a minimum duration for exclusion from the cohort. We have also included a location category to inform us if the resident made a local or longer distance migration. The exclusion of the residents who were present in the cohort for less than 80% of the year in 2008 reduced the cohort to 17,349.

The cohort will be followed continuously using routine health data (demographics, mortality, hospital admissions, and general practitioner records, including prescriptions) with periodic updates of housing regeneration intervention data. We will compare the health of residents within the homes before and after the housing regeneration work has taken place. Linking individual level health data to housing intervention data will allow us to estimate potential reductions in health service use costs as a direct result of housing regeneration. Routinely collected demographic data have allowed us to monitor cohort migration to obtain precise individual-level risk duration as a denominator. Routinely collected data from before, during and post intervention periods have also allowed us to calculate and control for potential confounders while retaining the maximum number of people in our intervention cohort.

Sarah E. Rodgers, Martin Heaven, Arron Lacey, Wouter Poortinga, Frank D. Dunstan, Kerina H. Jones, Stephen R. Palmer, Ceri J. Phillips, Robert Smith, Ann John, Gwyneth A. Davies, Ronan A. Lyons, Cohort Profile: The Housing Regeneration and Health Study. International Journal of Epidemiology 2012; 1-9, doi:10.1093/ije/dys200

Corresponding author email: s.e.rodgers@swansea.ac.uk

Utilising health informatics to detect an unintended consequence of antibiotics policy change: Increased rates of Acute Kidney Injury (AKI) post-operatively

Vadiveloo, T, Population Health Sciences Division, University of Dundee
Donnan, P, Population Health Sciences Division, University of Dundee
Bell, S, Renal Unit, Ninewells Hospital, Dundee
Patton, A, Scottish Antimicrobial Prescribing Group, Scottish Medicines Consortium, Glasgow
Sneddon, J, Scottish Antimicrobial Prescribing Group, Scottish Medicines Consortium, Glasgow
Marwick, C, Population Health Sciences Division, University of Dundee, Dundee
Bennie, M, National Medicines Utilisation Unit, Information Services Division, Edinburgh
Davey, P, Population Health Sciences Division, University of Dundee, Dundee

Introduction: Scottish antibiotic policies were revised in October 2008 due to an increase in Clostridium difficile infection (CDI). In patients undergoing orthopaedic surgery, recommended antibiotic prophylaxis changed from cefuroxime to flucloxacillin and gentamicin. However, the change in antibiotic policy has raised concerns about a possible increase in post-operative acute kidney injury (AKI).

Aim: To examine the rates of AKI among patients undergoing orthopaedic surgery in Tayside before and after the antibiotic policy change using record-linking detailed electronic healthcare data.

Methods: Patients who underwent orthopaedic surgery between the 1st October 2006 and 30th September 2010 were included in the study. Data linkage was performed using anonymised record linkage via the Health Informatics Centre (HIC). The post-operative renal impairment was defined using Acute Kidney Injury Network (AKIN) criteria from electronically captured laboratory data. The monthly rates of AKI, stratified into three levels of severity (Stages 1-3) were calculated and interrupted time series analysis of intervention effect was carried out in the two years before and after the change in antibiotic policy in October 2008.

Results: There were 7698 relevant orthopaedic procedures during the study period which were included in the analysis, with 767 cases of AKI. Based on 200 operations performed every month, overall, there was an increase in AKI cases from 6.5% prior to policy change to 10.1% after. Stage 1 AKI increased from 4.3% to 7.6%, stage 2 increased from 0.4% to 1.1% and stage 3 increased from 0.2% to 1% following the policy change. There was a significant increase in slope for the 48-hour change in creatinine after introduction of the policy (p=0.035), adjusted for gender, age and use of nephrotoxic drugs. On analysis of patients with fractured neck of femur group who received co-amoxiclav as prophylaxis throughout the study period, there was no significant change in slope for before compared to after intervention.

Conclusion: Extraction and analysis of routine electronic health data demonstrated an unintended consequence of an antibiotic policy change. The change in antibiotic policy was associated with significant increase in AKI in patients who underwent orthopaedic implant surgery within Tayside. Most of the increase in AKI was Stage 1 and transient. However, there was a significant increase in Stages 2 and 3, which persisted over the 7 day post-operative period. This has led to a further change in policy with co-amoxiclav now being given as prophylaxis to all patients undergoing orthopaedic surgery.

Corresponding author email: t.vadiveloo@dundee.ac.uk

Unintended consequences of change to antibiotic policy for patients with sepsis

Patton, A, Scottish Medicines Consortium, Healthcare Improvement Scotland
Davey, P, University of Dundee
Marwick, C, University of Dundee / NHS Tayside
Sneddon, J, Scottish Medicines Consortium, Healthcare Improvement Scotland
Nathwani, D, NHS Tayside

In 2008, the Scottish Antimicrobial Prescribing Group (SAPG) recommended changes in antibiotic policy [1] as one of several interventions to address increasing rates of Clostridium difficile infection (CDI) which is associated with significant morbidity and mortality. By restricting the use of antibiotics associated with a higher risk of CDI, the intended consequences were a reduction in the use of these antibiotics accompanied by a reduction in the CDI rate. SAPG required reassurance that patients still received adequate antibiotic therapy to effectively treat infections. Therefore a potential unintended consequence of an increase in 30-day mortality (from admission) was evaluated.

Interrupted time series analysis with segmented autoregressive error models was used to quantify intervention effects at fixed time points following the introduction of the restrictive antibiotic policy in one pilot board, NHS Tayside. Hospital admissions data from the local patient administration system were linked to the national register of deaths. Antibiotic usage data for acute admission units and and surgical wards were obtained from the local pharmacy store system and ward level CDI data were obtained from the microbiology laboratory system for patients admitted through an acute admission unit or to one of six surgical wards.

The preliminary results were presented at the International Forum on Quality and Safety in Healthcare conference in 2011. In medicine, there was a statistically significant 52% reduction in the use of restricted antibiotics, a 25% reduction in CDI and a non-significant 11% reduction in 30-day mortality (from admission) six months after the policy change. However, all-cause mortality was highlighted as not being sensitive enough to detect any changes in mortality in patients with infection.

Identifying patients with infection using ICD-10 codes is problematic as there is no specific code for “sepsis” and codes provide no method of indicating severity of infection so an alternative proxy marker for sepsis was required. The prevalence of sepsis was 60% in 2157 patients who had blood cultures taken in Ninewells. We compared mortality in all blood culture cases with 5839 patients who were in the same wards with similar length of stay. After adjustment for age, gender and co-morbidity the odds ratio for 30-day mortality in blood culture cases vs comparators was 3.97 (95% CI 3.34-4.72). Hospital admissions data, linked to the national registry of deaths, were linked to the local laboratory system to identify patients that had a blood culture taken. 30-day mortality (from date of blood culture) was calculated as a proxy marker for patients with sepsis. We found there was a non-significant reduction in mortality by 23% six months following the policy change. SAPG are currently investigating the feasibility of linking routine hospital admissions and data from the National registry of deaths with microbiology data held in Scottish Care Information (SCI) Store to evaluate positive and negative consequences of change to antibiotic policy across Scotland.

1. Chief Executive Letter (CEL) 30, July 2008
http://www.scottishmedicines.org.uk/files/sapg/CEL_30_2008_-_Antimicrobial_Prescribing.pdf

Corresponding author email: andrea.patton@nhs.net

Tools to support clinical care in Chronic Kidney Disease: Exploiting existing data through data linkage

Marks, A, University of Aberdeen and NHS Grampian
Prescott, GJ, University of Aberdeen
Smith, WCS, University of Aberdeen
Robertson, L, University of Aberdeen
Simpson, WG, NHS Grampian
Fluck, N, NHS Grampian
Black, C, University of Aberdeen and NHS Grampian

Background and Aims

A number of those with the common condition chronic kidney disease (CKD) will go on to require renal replacement therapy (RRT: dialysis or transplantation), however, not all will. For some there will be no further deterioration in kidney function. Not all of the 8-10% of the population with CKD therefore need to be seen by kidney specialists. The aim of this study was to use data-linkage of routine health care data to identify baseline characteristics that predict outcomes and then develop tools that, for a given individual, could predict likely outcome in the future.

Methods

A cohort with CKD in 2003 (GLOMMS-I) had previously been identified with measures of kidney function and baseline comorbidity (n~3400). Linkage to GRO and the local renal management system gave outcome measures for mortality and RRT. Models were developed to predict likely mortality and RRT outcomes at five years based on baseline characteristics. The GLOMMS-II cohort (a population based cohort of all with CKD in the region in 2003, 20,000 sample of those with normal kidney function and 20,000 sample of those with no measure of kidney function in 2003) had data-linkage to GRO, SMR01 records, Scottish Renal Registry and the local renal management system to generate a dataset with baseline characteristics and outcomes. The predictors of outcome were again reported and the utility of the prediction models developed in GLOMMS-I explored.

Results

Male sex, older age, more advanced kidney disease and the presence of worse proteinuria were associated with mortality in both GLOMMS-I and GLOMMS-II. However mortality prediction models that included measures of proteinuria and kidney function with age and sex performed no better than those with just age and sex alone so were not developed further. Male sex, more advanced kidney disease and worse levels of proteinuria were associated with the initiation of RRT. There was also a non-significant higher risk at younger age. Of the models developed a logistic regression model to predict RRT outcome at five years performed best in GLOMMS-I, c-statistic 0.938. When trialled in GLOMMS-II, the c-statistic was 0.958, and the performance using a 5% risk threshold (i.e. individual labelled as high-risk if predicted probability of initiated RRT by five years $\geqslant$ 5%) in GLOMMS-II was less sensitive (0.735 vs 0.824), but more specific (0.979 vs 0.901) than in the GLOMMS-I cohort. Overall 98% of people were correctly classified.

Conclusion

Data-linkage of routine health-care data offers a means to develop tools to guide the stratification of patient care in routine health care practice.

Corresponding author email: a.marks@abdn.ac.uk

Using Record Linkage of Routine Health Data to address Lithium Renal Safety: Analysis of longitudinal data in a Random Coefficient Model with estimated Glomerular Filtration Rate (eGFR)

CLos, S, Stratheden Hospital, Cupar; NHS Fife
Rauchhaus, P, HIC Dundee
Severn, A, Ninewells Hospital, Dundee; NHS Tayside
Donnan, P, Dundee Epidemiology and Biostatistics Unit, University of Dundee

Background: For over three decades there has been a debate about the effect of Lithium therapy on renal function. Two recent meta-analyses [Paul 2010; McKnight 2012] pointed out the poor quality of available study data and the need for epidemiological studies that control for confounders. Recent population based studies (Bendz 2010; Bocchetta 2013) did not provide an answer to this debate due to limitations of the cross-sectional study design (with associated problems as incidence-prevalence bias) and did not adjust for co-morbidities, co-prescribed medication and episodes of Lithium toxicity. With the CKD-EPI equation [Levey 2010] there is now a simple tool available to estimate GFR reliably above 60 ml/min/1.73m2, whilst the still widely used MDRD formula [Levey 1999] is associated with inaccuracy and negative bias.

Research Question: What is the annual decline in renal function in the Lithium exposed group and in the group exposed to comparator drugs?

Design: The design was a cohort study of patients newly commenced on maintenance treatment of lithium / comparator drug in Tayside between 2000 2011.
Primary outcome was the estimated Glomerular Filtration Rate (eGFR) using the CKD-EPI equation.

Data Sources: HIC provided deterministically electronically linked records from the following sources:

- HIC Prescribing database; raw data included 221,403 prescription records of index or comparator drugs from 6.892 patients, and 1,651,566 prescriptions of selected concomitant drugs.
- Scottish Morbidity Records (SMR01 - general admissions, SMR04 - psychiatric admissions)
- Death registry from the General Registry Office
- Scottish Indices of Multiple Deprivation deciles for social deprivation; these were included in a propensity score (to be treated with index drug or comparator drugs)
- Regional biochemistry

Statistical Analysis: The approach of choice was a random coefficients model (proc mixed; SAS 9.2) for repeated measures.

Results: 1120 patients aged between 18 and 65 years qualified for analysis. Adjusted for age, sex and baseline eGFR the estimated mean annual decline in eGFR was 1.5 ml/min/1.73m2 (SE 4.2 ml/min/1.73m2) for both groups. The random coefficients model identified predictors for a decline as comorbidities (diabetes and hypertension) and co-prescriptions of nephrotoxic drugs, but not length of exposure to Lithium.

Conclusions: Health Informatics provided via deterministic linkage of routine electronic data an ideal opportunity to address a question in drug safety, overarching the fields of Psychiatry, Nephrology and Biochemistry. Furthermore, the rich data environment of HIC enabled us to build a model to identify predictors for a decline in renal function.

Corresponding author email: stefan.clos@nhs.net

Developing a virtual population based cohort to study Chronic Kidney Disease - GLOMMS-II (the second Grampian Laboratory Outcomes Morbidity and Mortality Study)

Marks, A, University of Aberdeen and NHS Grampian
Prescott, GJ, University of Aberdeen
Smith, WCS, University of Aberdeen
Robertson, L, University of Aberdeen
Simpson, WG, NHS Grampian
Fluck, N, NHS Grampian
Black, C, University of Aberdeen and NHS Grampian

Background and Aims

Chronic kidney disease (CKD) is common, important and associated with the need for renal replacement therapy (dialysis or transplantation). However, with the advent of routine reporting of eGFR (a measure of renal function), the lack of knowledge of outcome in those with less advanced kidney disease has been highlighted. The aim of the GLOMMS-II study was to identify a cohort across the range of kidney impairment and comparison groups to study risk factors for poor outcomes in CKD.

Methods

All measurements of excretory renal function in the region from 1999-2009 were available. All those with abnormal renal function in 2003 (whether due to CKD or other entity), a 20,000 sample of those with normal kidney function and a 20,000 sample of those without a measure of kidney function in 2003 were included. Data-linkage to hospital episode records (SMR01) in the five years prior to 2003 allowed for assessment of baseline comorbidity. Mortality and other outcomes including renal replacement requirements were assessed through data-linkage within the Grampian Data Safe Haven to data from the National death registry, SMR01, Scottish Renal Registry and local renal management systems. A previously identified subset of ~3000 with advanced CKD (GLOMMS-I) and casenote review were used to explore the validity of the SMR01 based comorbidity with kappa for agreement, sensitivity and specificity.

Results

At baseline, those with stage 3-5 CKD were more likely to be female and older than those who had normal kidney function. Those with CKD were less likely to live in the least deprived areas than those with normal kidney function. Amongst those comorbidities associated with the development of CKD including vascular disease (ischaemic heart disease, congestive cardiac failure, peripheral vascular disease, hypertension), diabetes and connective tissue disease the odds of these comorbidities being present at baseline increased with worse level of kidney function. The age-sex corrected odds of having diabetes for stage 3a (moderate) CKD were 3.6 (3.1-4.1) (vs normal) and 12.4 (8.0-19.1) for stage 5 CKD (the worst). Performance of SMR01 compared to casenote review was at least moderate if not substantial or good, specificity was generally above 90%, although sensitivity was less good.

Conclusion

The use of routine health-care data offers a robust means of constructing disease cohorts with similar control groups to study outcome in chronic disease, by assessing the baseline characteristics in a standard and reproducible way in both cases and controls.

Corresponding author email: a.marks@abdn.ac.uk

Comparing Relational and Graph Databases for Pedigree Datasets

Kirby, G, University of St Andrews
Kerckhove, C, unknown
Shumailov, I, unknown
Carson, J, unknown
A, No, Dibben
C, No, Williamson
L, No,

Increasingly large family pedigree datasets are being constructed from routine civil and religious registration data in various parts of the world. These are then being used in health, social and genetic research in a variety of different ways. Often the type of questions that are being asked involve complex queries, such as the degree of relatedness between multiple sets of individuals, and involve traversing through the data typically multiple times. There is therefore an important issue of efficiency of querying. In this paper we evaluate the suitability of two classes of database, relational (MariaDB) and graph (Neo4J), for storing and querying pedigree datasets representing millions of individuals. We report results of measurements of scalability, query performance, and the ease of query expression, performed on both synthetic and real datasets.

Corresponding author email: graham.kirby@st-andrews.ac.uk

Using linked data to identify potential bias due to missing paternal details in birth registrations

Sims, S, Telethon Institute for Child Health Research
O'Donnell, M, Telethon Institute for Child Health Research

Fathers play an important role in children's lives and their impact on child health outcomes. However research predominantly has focused on the mother's role with fathers often neglected. Families with fathers not living at home are more likely to be faced with financial disadvantage, include both parents who have health problems, are relatively young and have low education levels. To obtain information on fathers, particularly for linked population level data, birth registrations are often used, however sometimes fathers information is missing.

The aim of our study was to determine how much missing paternal data is in birth registrations and the characteristics of families that were likely to have missing data. We hypothesised that families with a high risk of adverse health outcomes were more likely to have missing paternal details on birth registrations. This study used linked Western Australian population level data including Birth Registration, Midwives Notification data, and Child Protection data.

Results from our study showed that during 1980 - 2005 there was an average of at least 4.4% (27,913 cases) of birth registrations with missing father's data. The prevalence of missing paternal data has changed over that time but the characteristics of these families have not changed. In particular there was a significant decreasing trend of 2.6% cases per year where father's data was missing. Some of the characteristics of families most at risk of missing paternal information included those who lived in areas of high socio-economic disadvantage, had young single mothers, a child and mother of Aboriginal origin, smoked during pregnancy, had a short gestation period, and low birth weight.

The strong association between adverse child health outcomes and the absence of paternal information might be a source of bias in existing data. This bias may be introduced during the analysis of data because records are often excluded due to missing information. Excluding records with missing paternal information (e.g. age, Indigenous status) could actually lead to families that are at highest risk of adverse outcomes being removed. This would result in a biased sample rather than a representative cross-section of the population. If this sample were to be used in research subsequent results could be misleading. Therefore researchers need to ensure that, wherever possible, families with missing father's data remain in the sample to improve accuracy of research findings. It is also highlights that fathers make an important contribution to children's outcomes and are an essential part of research in this area.

Corresponding author email: ssims@ichr.uwa.edu.au

Generation of Family Units in the SAIL Databank

Tingay, K.S, College of Medicine, Swansea University

Family relationships, both biological and environmental, have long been the subject of health and social care research. The ability to group individuals by family unit has clarified the aetiology of many inherited medical and psychological conditions, and allowed researchers to monitor the spread and control of diseases. Sociologically, too, family studies have provided many important insights, including the burden of caring for other family members, and the impact of being placed in care. Being able to link family unit members could also be useful when creating control groups, as the researcher would be able to flag those who are exposed to a similar environment to the subject(s) of research. The ability to examine individuals who share a close common environment, therefore, has huge potential, especially when coupled with large datasets.

The Secure Anonymous Information Linkage databank (SAIL) holds routinely collected national health data and social care datasets including hospitals, clinics and GP practices in Wales. Data sources are not restricted to health and include other settings, such as social services, housing, transport and education. Patients are linked through different datasets using an Anonymised Linkage Field (ALF). In 2009, SAIL introduced the Residential Anonymised Record Linkage field (RALF) which allows an individual to be placed in relation to a LSOA while still ensuring anonyminity. While this allows individuals to be placed in a geographical context, there had not previously been attempts to infer family units living in the same RALF.

The aim of the project is to develop a method to infer family units consisting of multiple ALFs living within and moving through RALFs. This model must be able to track anonymised individuals within an inferred family unit as they move through different households, map changes to households over time (as members enter and leave the household), and to flag life events which may impact on the physical and/or emotional wellbeing of the household members. We propose to use this dataset to write and test an algorithm to predict Mother-Child relationships.

The results will be tested on a cohort of family units. Issues encountered will be discussed in relation to a set of fictional case study families consisting of 16 individuals moving over 8 different addresses over the course of several years. The presentation will address how households have been identified in SAIL, how changes to families and households were addressed, limitations to the project, and future work.

Corresponding author email: k.s.tingay@swansea.ac.uk

Blood pressure and the initial presentation of twelve cardiovascular diseases: a CALIBER study

Rapsomaniki, E, Clinical Epidemiology Group, Department of Epidemiology and Public Health, UCL
Pujades-Rodriguez, M, Clinical Epidemiology Group, Department of Epidemiology and Public Health, UCL
George, J, Clinical Epidemiology Group, Department of Epidemiology and Public Health, UCL
Shah, A, Clinical Epidemiology Group, Department of Epidemiology and Public Health, UCL
Denaxas, S, Clinical Epidemiology Group, Department of Epidemiology and Public Health, UCL
Smeeth, L, Department of Non-communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine
Timmis, A, National Institute for Health Research, Biomedical Research Unit, Barts Health London
Hemingway, H, Clinical Epidemiology Group, Department of Epidemiology and Public Health, UCL

Background

It is not known how blood pressure (BP) is associated with the initial presentation of twelve of the most common cardiovascular diseases (CVDs). Previous bespoke, investigator led studies had limited clinical resolution because of small size, narrow range of CVDs studied or both. No previous large-scale studies using linked electronic health records have address this question.

Objective

To determine whether systolic BP (SBP) and diastolic BP (DBP) differ in their associations with a wide range of initial CVD presentations.

Methods

Cohort analysis of linked electronic health records (Clinical Practice Research Datalink, Hospital Episodes Statistics, Myocardial Ischaemia National Audit Project and Office for National Statistics) from 1.93 million patients aged ⩾ 30 years and with no prior CVD. Study endpoints were stable and unstable angina, myocardial infarction (MI), coronary death, heart failure, ventricular arrhythmia, transient ischaemic attack, ischaemic stroke, intracerebral and sub-arachnoid haemorrhage, abdominal aortic aneurysm (AAA) and peripheral arterial disease (PAD). Cox proportional hazard regression was used on multiply-imputed data to estimate associations between SBP and DPB and each CVD endpoint. Effects were estimated per 1 standard deviation increase in blood pressure and for 10 mmHg-categories of SBP and DBP.

Results

A total of 11.6m person-years of follow-up and 103,130 CVD events were analysed. For each CVD endpoint the risk increased with higher BP level down to 90 mmHg SBP and 70 mmHg DBP. SBP was most strongly associated with intracerebral haemorrhage (HR=1.36, 95%CI 1.31-1.42) and stable angina (HR=1.32, 95%CI 1.29-1.35); weakly associated with cardiac arrest or sudden cardiac death (HR=1.13, 95%CI 1.08-1.19) and not associated with AAA (HR=1.01, 95%ci 0.96-1.07). DBP had stronger association with AAA (HR=1.22, 95%CI 1.14-1.30), no association with PAD (HR=1.01, 95%CI 0.99-1.04) and only weak to moderate associations with other CVD outcomes. Compared to DBP, SBP showed stronger associations with stable angina, PAD, ischaemic stroke, MI and heart failure. Effects declined with age, especially for stable angina, unheralded coronary death and heart failure; with steeper age-related declines observed for SBP.

Conclusions

A BP higher than 90mmHg systolic and 70mmHg diastolic was linearly associated with increased risk of most clinical phenotypes of CVD, but not all. Marked heterogeneity in the strength of the association was observed according to disease type, and the strength of the associations declined with older age.

Corresponding author email: h.hemingway@ucl.ac.uk

Prevalence, management and healthcare burden of irritable bowel syndrome in Scotland

McTaggart, SA, ISD Scotland, NHS National Services Scotland
Wyper, G, ISD Scotland, NHS National Services Scotland
Harkins, L, ISD Scotland, NHS National Services Scotland
Bishop, I, ISD Scotland, NHS National Services Scotland
Bennie, M, ISD Scotland, NHS National Services Scotland and University of Strathclyde

METHODS: A consultation records database covering 56 general practitioner practices in Scotland (~255,000 people; 5.8% of the population) was used to identify consultations primarily due to IBS between April 2009 and March 2011 (read codes: IBS, constipation, diarrhoea). Read codes suggesting other causes of diarrhoea or constipation and patients with inflammatory bowel disease, diverticular disease, coeliac disease or bowel cancer were excluded. Prescriptions for antispasmodics, laxatives or antidiarrhoeals, referrals to outpatient clinics and acute hospital admissions between January and December 2011 were also analysed using national datasets.

RESULTS: Based on consultation records, an estimated 341,180 adults (⩾ 18 years) in Scotland suffer from IBS, representing an estimated prevalence of 7.7% (women: 9.8%; men: 5.5%). During 2011, 142,738 adults received ⩾ 1 prescription for antispasmodics, most frequently mebeverine (40.1%), hysocine butylbromide (35.7%) and peppermint (18.0%), giving an estimated prevalence of antispasmodic-treated IBS of 3.4% (women: 4.7%; men: 2.0%). Many (33.6%) of the patients receiving antispasmodics were also prescribed laxatives (24.5%), the antidiarrhoeal loperamide (6.7%) or both (2.5%). Of the patients treated with antispasmodics, 11,645 (9.0%) visited a gastroenterology outpatient clinic in 2011 (11.7% of all gastroenterology clinic attendances) and 1,869 (1.3%) were acutely admitted to a Scottish hospital due to IBS or symptoms likely to be associated with IBS, most frequently constipation (80.3%). The average length of hospital stay was 2.1 days (2.4 days for admissions due to constipation).

CONCLUSIONS: The estimated prevalence of IBS in Scotland is 7.7% based on consultation records and 3.4% based on prescription records. These estimates do not include patients who use over-the-counter medication or do not consult their GP. IBS accounts for a significant proportion of Scottish gastroenterology workload. Better understanding of the use of medicines and health services in the management of IBS will help to support improved resource use and patient care.

NOTE: This study was undertaken as a commercially contracted piece of work for Almirall UK

Corresponding author email: stuart.mctaggart@nhs.net

Prevalence and treatment of Active Asthma in Scotland using the Prescribing Information System

Steiner, M, University of Aberdeen
Devereux, G, University of Aberdeen
Turner, S, University of Aberdeen
McLay, J, University of Aberdeen
Bishop, I, ISD National Services Scotland
Wyper, G, ISD National Services Scotland

Background

Many studies of the clinical epidemiology of asthma that have used routinely collected drug prescription or dispensing data have been limited to samples assumed to be representative of the national population from which they are drawn. Our aim was to describe asthma prevalence and treatment in children and young adults using the Prescribing Information System (PIS), a national prescribing and dispensing database for Scotland.

Methods

For more than 95% of the dispensed prescriptions in primary care from December 2009 a valid patient identifier is available and the database includes some socio-demographical characteristics (age-group, sex, SIMD) of the patients. Data were also linked to hospital admission data. The analysis was limited to patients aged up to 50 years to reduce contamination by COPD. For this analysis the data are restricted to 44 years to match practice population denominators.

Data

We identified 358,804 patients with 2,809,563 dispensed prescriptions for inhaled therapies used for asthma; equating to a prevalence of 11.4% of the 3,139,356 people aged 0-44 registered with a GP in Scotland. The age specific prevalence rates are detailed in the table. However, 95,207 patients had only one or two dispensed prescriptions for short-acting beta2-agonists (SABA) and no other inhaled therapies in the two years; we consider these patients to be unlikely to have active asthma (table). Additionally, 1,041 cases on inhaled therapy had hospital admission(s) with a diagnosis of COPD (ICD10: J40-J44) and are excluded from further analysis. 6,056 (2.3%) of people collecting inhaled therapy (>2 SABA) had at least one hospital admission with a primary diagnosis of asthma. We used the stepwise pharmacological management from the BTS/SIGN guidelines to classify the patients into the treatment step according their dispensing history.

18.4% of patients collected SABA only, 46.8% collected SABA + inhaled corticosteroids (ICS), 0.1% (371) collected SABA +long acting beta2 agonist (LABA) only, 13.6% collected SABA +combined ICS/LABA preparation, 2.1% collected SABA + ICS +LABA, 1.7% collected ICS/LABA only, leukotriene receptor antagonists (LTRA) were collected by 8.1% and long acting antimuscarinic agents (LAMA) were collected by 1.0%. The classification into the management steps is currently under way.

Conclusion

This current and whole population database indicates that the prevalence of asthma is approximately 10% in young adults and 15 % in children living in Scotland but prevalence of active asthma is approximately 8% in adults and 10% in children. The extension of the dataset for another year (till December 2012) in the near future should allow some refinement of our results presented so far.

Corresponding author email: m.steiner@abdn.ac.uk

Exploiting record-linkage to alcohol-related hospitalisation and mortality data to quantify non-response bias in the Scottish Health Survey

Gorman, E.L, MRC/CSO Social and Public Health Sciences Unit, Glasgow, UK
Leyland, A.H, MRC/CSO Social and Public Health Sciences Unit, Glasgow, UK
McCartney, G, NHS Health Scotland, Glasgow, UK
White, I.R, MRC Biostatistics Unit, Cambridge, UK
Katikireddi , S.V, MRC/CSO Social and Public Health Sciences Unit, Glasgow, UK
Rutherford, L, ScotCen Social Research, Edinburgh, UK
Graham , L, ISD, NHS NSS, Edinburgh, UK
Gray, L, MRC/CSO Social and Public Health Sciences Unit, Glasgow, UK

Background

National health surveys are a widely utilised resource for monitoring population health, but the validity of inference relies upon survey representativeness of the population of interest. Many surveys are increasingly subject to high rates of non-response, which may lead to non-response bias if non-respondents differ systematically from respondents. Differences in health-related behaviours may be indicated by differential health outcomes in follow-up, such as alcohol-related hospital admission or mortality. A dearth of reliable health-related information allowing comparison between non-responders and responders can present a barrier to identifying and adjusting for these differences. The aim of this study is to exploit confidentially record-linked hospital admission and mortality data to assess the representativeness of the Scottish Health Survey in terms of alcohol-related harms.

Methods

The cross-sectional Scottish Health Surveys are designed to provide estimates representative of the Scottish population. The 2003 SHeS achieved a 67% response rate; for the 91% who consented survey records were linked to routine hospital admission and mortality data (~90% accurate diagnosis, 99% complete), with linkage conducted up to the end of 2011. Population, hospital and mortality data for the general population of Scotland are also available. We compared directly age-standardised survey weighted estimates of alcohol-related harms - alcohol-related mortality or at least one alcohol-related hospital admission - in the 2608 men and 3330 women aged 20-69 years at interview in SHeS 2003-SMR/NRS to those in the population of Scotland using contemporaneous data.

Results

Among men in the SHeS study sample, 92 (3.5%) were hospitalised and 9 (0.35%) died from alcohol-related causes. The corresponding figures for women were 64 (1.9%) and 3 (0.10%), respectively. Alcohol-related harms among men were significantly lower in the SHeS sample (414 per 100,000 person-years [95% CI: 342-493]) relative to the Scottish population (680 [676-685]), with a rate ratio of 0.61 (0.50-0.73). For women, alcohol-related harms were lower in the SHeS sample (236 [183-305]) compared with the Scottish population (277 [274-280]), with a corresponding rate ratio of 0.85 [0.66-1.10]).

Conclusions

After applying standard socio-demographic weighting adjustments and age-standardisation we find lower alcohol-related harm among 20-69 year old 2003 SHeS respondents relative to the general population of Scotland, suggesting there may also be a corresponding differential in alcohol consumption. Record-linked hospital and mortality records offer a rich set of information to explore and address health-related non-response bias.

Corresponding author email: e.gorman@sphsu.mrc.ac.uk

What can linkage to electronic patient records tell us about differences in help-seeking behaviour and health-related outcomes in relation to participation in observational studies?

Cornish, RP, School of Social and Community Medicine, University of Bristol
Boyd, AW, School of Social and Community Medicine, University of Bristol
Van Staa, T, CPRD, MHRA
Macleod, J, School of Social and Community Medicine, University of Bristol

Background

One of the key strengths of longitudinal studies is their ability to investigate causality by measuring exposures and outcomes at multiple time points. However, attrition is unavoidable in large longitudinal studies and this can lead to biased results. Linkage to health and administrative data provides a way of determining the likely extent of this bias by providing data on those who did and did not participate in a research study.

Methods

The Avon Longitudinal Study of Parents and Children (ALSPAC) is a large birth cohort with detailed data collected from before birth through to late adolescence. Approximately 14,000 pregnant women were recruited into the study (out of around 20,000 who were eligible) between 1991 and 1992. Data collection continues, primarily via questionnaires and clinics, but also through linkage to routine data. The NHS Information Centre (NHS IC) linked subjects eligible to participate in ALSPAC to the General Practice Research Database (GPRD), an anonymised database of primary care records of around 5 million patients in the UK. [GPRD has now been incorporated into the Clinical Practice Research Datalink (CPRD)]. Hospitalisation rates, using HES data linked by CPRD, GP consultation rates and prescription rates were defined in 5-year age bands. In addition, subjects were classified according to whether they had ever smoked, ever been pregnant and ever had an "at risk" Read code entered in their GP record. Using ALSPAC's administrative data, subjects were classified according to whether or not they had ever participated in the study and whether or not they had participated recently; they were also put into groups according to their participation score, a composite score taking into account their overall participation in terms of both clinic attendance and completion of questionnaires. The data were analysed in a safe setting at the CPRD offices. Logistic regression was used to analyse the binary outcomes; consultation, hospitalisation and prescription rates were analysed using multilevel Poisson models.

Results

Preliminary analyses suggest that those with higher levels of participation, particularly recent participation, in ALSPAC are different in several ways to those with no recent participation. Data analysis is ongoing and final results will be presented.

Conclusions

As well as considering the implications of the results of this analysis, the limitations of this particular linkage and ways in which these can be overcome will be discussed.

Corresponding author email: rosie.cornish@bristol.ac.uk

Accuracy of electronic health record data for ascertainment and sub-classification of stroke outcomes in large-scale epidemiological studies: a systematic review from the UK Biobank stroke outcomes group

Woodfield, RM, University of Edinburgh
Grant, I, Information and Statistics Division, National Services Scotland
Sudlow, C, University of Edinburgh and UK Biobank

Introduction:
UK Biobank is a very large prospective study of 500,000 middle-aged participants, recruited in England, Scotland and Wales between 2006 and 2010. Long-term follow-up includes cohort-wide linkages to National Health Service death and cancer registries, hospital discharge data and primary care data. To help inform UK Biobank's approach to the ascertainment, confirmation and sub-classification of stroke outcomes, we performed a systematic review of the accuracy of such data for stroke and its main pathological types (ischaemic stroke, intracerebral haemorrhage, and subarachnoid haemorrhage).

Methods:
We comprehensively sought studies of the diagnostic accuracy of coded electronic health record data for stroke and its main pathological types. We included studies published from 1990-2012, comparing data from death certificates, hospital discharge records or primary care records versus a reference standard for stroke and/or its pathological types. From each study we extracted data on the study setting and subjects, data source(s), cerebrovascular codes(s) validated, reference standard used, and - where available - the positive predictive value (PPV) and sensitivity of the code(s). We also obtained and analysed data on International Classification of Disease (ICD) stroke codes in UK national hospital discharge records.

Results:
From >5500 articles identified by our search, we included 38 studies. Most were of hospital discharge records, coded using ICD version 9 or 10. These studies were heterogeneous, with substantial variation in the accuracy of the cerebrovascular code(s) validated: PPV for stroke of any pathological type ranged from 31-97% (27 studies), increasing with use of fewer, more specific stroke codes and with restriction to codes in the primary diagnosis position only. Sensitivity ranged from 66-96% (8 studies), increasing with use of broader cerebrovascular disease codes and with inclusion of relevant codes in both primary and secondary positions. In UK hospital discharge records, around a quarter of stroke codes in recent years were in the "unspecified stroke" category but this proportion is declining. Only two studies assessed electronic primary care data, and only one assessed use of several overlapping data sources versus an independent reference standard.

Conclusions:
Selecting the most appropriate combinations of codes from hospital discharge records, together with the declining use of less informative codes, should improve the accuracy of ascertainment and confirmation of stroke outcomes in large epidemiological studies such as UK Biobank. Further UK-based studies are required to assess the added benefits of combining several overlapping data sources, including primary care data.

Corresponding author email: rebecca.woodfield@ed.ac.uk

Vascular disease in women: Comparison of diagnoses in hospital episode statistics and general practice records in England

Wright, FL, University of Oxford
Green, J, University of Oxford
Canoy, D, University of Oxford
Cairns, BJ, University of Oxford
Balkwill, A, University of Oxford
Beral, V, University of Oxford

Background

Electronic linkage to routine administrative datasets, such as the Hospital Episode Statistics (HES) in England, is increasingly used in medical research. Relatively little is known about the reliability of HES diagnostic information for epidemiological studies. In the United Kingdom (UK), general practitioners hold comprehensive records for individuals relating to their primary, secondary and tertiary care. For a random sample of participants in a large UK cohort, we compared vascular disease diagnoses in HES and general practice records to assess agreement between the two sources.

Methods

Million Women Study participants with a HES record of hospital admission with vascular disease (ischaemic heart disease [ICD-10 codes I20-I25], cerebrovascular disease [G45, I60-I69] or venous thromboembolism [I26, I80-I82]) between April 1st 1997 and March 31st 2005 were identified. In each broad diagnostic group, and in women with no such HES diagnoses, a random sample of about a thousand women was selected for study. We asked each woman's general practitioner to provide information on her history of vascular disease and this information was compared with the HES diagnosis record.

Results

Over 90% of study forms sent to general practitioners were returned and 88% of these contained analysable data. For the vast majority of study participants for whom information was available, diagnostic information from general practice and HES records was consistent. Overall, for 93% of women with a HES diagnosis of vascular disease, general practice records agreed with the HES diagnosis; and for 97% of women with no HES diagnosis of vascular disease, the general practitioner had no record of a diagnosis of vascular disease. For severe vascular disease, including myocardial infarction (I21-22), stroke, both overall (I60-64) and by subtype, and pulmonary embolus (I26), HES records appeared to be both reliable and complete.

Conclusion

Hospital admission data in England provide diagnostic information for vascular disease of sufficient reliability for epidemiological analyses.

Reference:

Wright et al. BMC Medical Research Methodology 2012, 12:161

Corresponding author email: lucy.wright@ceu.ox.ac.uk

A case for matching without names - an assessment of a cross-sector health-education data linkage using limited identifiers

Clark, D, NHS National Services Scotland
King, A, Scottish Government, ScotXed
Wood, R, NHS National Services Scotland
Mackay, D, University of Glasgow
Pell, J, University of Glasgow

Introduction

One of the objectives of this study was to test the technical feasibility, completeness and accuracy of data linkage across two sectors, in the absence of names.

The primary datasets to be linked were:

- ScotXed - All primary, secondary and special school pupils.
- Community Health Index (CHI) - All people registered with a General Practice in Scotland.

Methods

An extract of national education data (pupil censuses 2006-2011) was provided by ScotXed for linkage to relevant health datasets by NHS National Services Scotland Information Services Division (ISD). ScotXed contained pupil identifiers (date of birth, gender, postcode) but not names. ISD linked these to CHI, adapting previously developed algorithms refined to optimise the linkage success. Annual pupil censuses were linked to CHI for single years (method-A) and also by internally linking the pupil records across all 6 years (method-B). Validation of the linkage was undertaken by:

1. Submitting names from the resultant CHI records back to ScotXed for cross-comparison.
2. Taking a subset of pupils who also appear on the Scottish Qualifications Authority (SQA) database and using the names available from this dataset to provide a robust linkage and compare the CHI numbers found using both methods.

Results

Links to CHI were partitioned into 13 categories (A-M) dependent on levels of agreement of the 3 matching items, and distance to the nearest rivalled match in terms of probabilistic weight. The precision of matching (proportion of positives which are correct) within these partitions were assessed and grouped further into “safe” and “non-safe” matches. The precision for the linkage of the 2011 census using method A was 97.5% (all matches) and 99.5% (safe matches) while the sensitivity (recall) was nearly 100% (all matches) and 95.2% (safe matches). Using method B, precision for all matches increased to 98.7% while sensitivity for safe matches increased to 96.7%. These quality metrics were confirmed by the SQA linkage.

Discussion

The lack of individual names within the ScotXed data is representative of the real challenges posed by cross-sectoral linkage to which solutions are required. The results of this quality assurance show that, for research purposes, education records can be easily and robustly linked to health without recourse to names, suggesting that this approach may also have applicability to other sectors. By limiting the range of personal identifiers required, this approach also reduces the perceived, or real, privacy risk associated with linkage and may have public acceptability benefits.

Corresponding author email: dclark5@nhs.net

How important are data-specific match-weights to probabilistic linkage?

Henderson, KL, Public Health England; Institute of Child Health, UCL
Minaji, M, Public Health England
Muller-Pebody, B, Public Health England
Wade, A, Institute of Child Health, UCL
Gilbert, R, Institute of Child Health, UCL

Introduction

In probabilistic linkage, the probability that common identifiers match is computed, based on counts of identifiers agreeing, disagreeing or missing, from a linked gold-standard dataset containing only true matches and true non-matches. These probabilities generate total match-weights that are used to link the full datasets, creating a hierarchy of links that ideally separate true from non-true matches.

Public Health England's (PHE) national laboratory surveillance database (LabBase2) uses established match-weights developed to link individuals within LabBase2, but their suitability for linking LabBase2 to other datasets has never been questioned. To assess whether PHE match-weights are optimal for linking paediatric LabBase2 to Hospital Episode Statistics (HES) data, we compared them with two "new" sets of match-weights calculated from linked, gold-standard LabBase2/HES paediatric data.

Methods

Two sets of match-weights were calculated from two datasets of linked LabBase2/HES reports of children (1m-18yrs old) with bloodstream infections (January 2009-March 2010):

Dataset 1) pairs matched on NHS number only;

Dataset 2) pairs matched using all identifiers (NHS number, hospital number, sex, date of birth, postcode) to account for matches where NHS number was missing. Matched pairs above a conservative threshold were selected by manual review.

We compared the percentage of identifier match-weight contributions towards the total match-weight for agreement, disagreement or missing, for each set using radar-plots.

Results

The new match-weights were calculated from 10,436 (Dataset 1) and 15,046 (Dataset 2) paediatric true-matches, compared to >1 million PHE matches from all age-groups. The proportional distribution of match-weights assigned to each identifier for "agreement" was similar for all three sets. In contrast, "disagreement" on date of birth or NHS number resulted in a proportionally higher negative weight (16-20% and 18-27% respectively) for both new match-weight sets compared to PHE weights (14% for both), whereas postcode and hospital number weightings were low for both new sets (3% for both) in comparison to PHE weights (8% and 14% respectively).

Conclusions

The comparison of the three match-weights illustrates how important correct date of birth and NHS number are to predicting a true match for linked LabBase2 and HES paediatric data compared to other variables identified by PHE match-weights alone. This difference between the match-weight sets demonstrates the value of considering the population of interest and identifiers contained in both linkage datasets, as this may ultimately impact on the final linked dataset selected for analysis.

Corresponding author email: katherine.henderson@phe.gov.uk

A generalisable method to enhance observational research through linkage to electronic primary care records.

Boyd, A, University of Bristol
Middleton, R, University of Swansea
Davies, A, University of Bristol
Cornish, R, University of Bristol
Thompson, S, University of Swansea
Lyons, R, University of Swansea
Ford, D, University of Swansea
Macleod, J, University of Bristol

Background:
Within the UK, approximately 90% of population contact with the NHS occurs in primary care and is recorded in general practice (GP) electronic patient records (EPRs). EPRs offer a cost-effective means of retrospective and prospective follow-up for longitudinal studies. While centralised EPR databases exist, these provide limited follow-up potential for population based studies because of low population coverage. We have developed a methodology to link study participants to their EPRs via automated data extraction from individual practice software systems.

Methods:
Through a postal campaign, consent to link to EPRs was sought from the index children of the Avon Longitudinal Study of Parents and Children (ALSPAC). A consent exemption was granted, on the recommendation of the National Information Governance Board (NIGB), where gaining participant consent was not practicable. A pilot sample of explicitly consenting participants was selected. The NHS Information Centre identified the current registered practice of each participant based on NHS number. A senior partner at each practice, the data controller of the EPRs, was asked to assent to their practice records being accessed. Over 80% of practices identified used either clinical software supplied by Egton Medical Systems or hosted audit software supplied by Apollo Medical Systems. In assenting practices, these suppliers extracted copies of the EPRs of ALSPAC participants into the Secure Anonymised Infrastructure for Linkage (SAIL) at the University of Swansea. Within SAIL the records were cleaned and anonymised using a split file approach. An anonymised linking field allowed subsequent matching to the relevant ALSPAC participant.

Results:
Of the ~14,000 participants asked to consent 27% responded. Of these, over 95% consented to linkage. A pilot sample of 2,834 consenting participants, registered at 533 practices, was selected. While we are still extending our coverage, to date 249 practices (47%) have authorised data extraction, 16 (3%) have declined and 268 (50%) of practices are yet to report a decision. We have successfully started to extract records; extraction will continue during 2013.

Conclusions:
The population coverage of centralised EPR databases is being expanded but the eventual coverage remains unknown. This methodology provides a secure means to link to EPRs of participants in a cohort study. This allows individual follow-up which appears acceptable to participants and practices. It is also compliant with all relevant information governance requirements. This method is generalisable and forms a low-cost and effective means for any study to establish a link to their participants' primary care records.

Corresponding author email: a.w.boyd@bristol.ac.uk

Privacy Preserving Probabilistic Record Linkage (P3RL) - results from a pilot study using cancer registry data

Spoerri, A, Institute of Social and Preventive Medicine, University of Bern, Switzerland
Schmidlin, K, Institute of Social and Preventive Medicine, University of Bern, Switzerland
Clough-Gorr, K, Institute of Social and Preventive Medicine, University of Bern, Switzerland

Objective: To assess feasibility and performance of Privacy Preserving Probabilistic Record Linkage (P3RL) with Bloom filter encryption by linking data from a regional cancer registry (CR) to the national Swiss Childhood Cancer Registry (SCCR)

Methods: The SCCR only registers cancers in persons up to 21 years old. The regional Swiss CRs register cancers for all ages. In order to identify second cancers that occurred in adult age persons with a previous childhood cancer SCCR data (N=8,803 records) were linked to a single regional CR (N = 75,939 records). . The SCCR and CR data used for record linkage was: first name, last name, date of birth, date of death, sex, nationality, marital status, and community of residence. To protect the case's privacy, personally identifiable information (names, date of birth and date of deaths) were encrypted using Bloom filters. Two linkages were performed: a) with encrypted names and b) encrypted names, date of birth and date of death. We used a set of P3RL tools for the automatic pre-processing and encrypting of variables. We adapted GRLS, the linkage software from Statistics Canada, to handle Bloom filter hash codes and calculate dice-coefficients. The SCCR was previously linked to the regional CR in 2010 when it was still possible to use un-protected personal information (e.g. plain names). We compared the outcome of the P3RL with the linkage with plain names.

Results: We could link 97.9% of the potential linkable SCCR records to the regional CR when encrypting names alone. The second linkage including encrypted names plus date of birth and date of death resulted in 97.2% of the potential linkable SCCR records being linked. Additionally we found 11 records in the regional cancer registry which could be second tumors from primarily recorded children in the SCCR. Comparing the original linkage with plain names to the new P3RL with encrypted names showed that 99.4% of the records could be linked using encryption.

Conclusion: Protecting privacy using P3RL seems a feasible and valid alternative to the use of unencrypted names in record linkage projects.

Corresponding author email: spoerri@ispm.unibe.ch

Handling Large Volumes of Routinely collected Data

DSILVA, R, Swansea University
Jones, C, Swansea University
Brooks, C, Swansea University
Thompson, S, Swansea University
Jones, K, Swansea University
Ford, D, Thompson

As data volumes in all areas have seen exponential growth in recent decades, it has brought new challenges when it comes to managing and analysing these large volumes of data. In this presentation, we will introduce the SAIL (Secure Anonymised Information Linkage) Databank, a large database environment, and the principles used in its development. The SAIL Databank comprise of over 6.9 billion rows of anonymised routine data spanning many linked datasets. We will describe the various database methodologies (data distribution across multiple servers, table partitioning, compression, data clustering, indexes etc.) used to develop, scale, administer and maintain a high performance database environment. We will also discuss other useful data management techniques, including methods for quick movement and transformation of data, efficient SQL construction, architecting data processing solutions that exploit parallelism.

Corresponding author email: r.dsilva@swansea.ac.uk

The Data Appliance - It is time to get data coming to us

Thompson, s, Swansea University, SAIL

The SAIL Databank is introducing numerous enterprise grade Data Appliance's into the NHS and other organisations to collect, catalogue and prepare dataset for research. The Data Appliance will enable these organisations to link their datasets, using the latest matching technologies, to provide a unique view for the organisation of their own data. The Data Appliance will make it easier for these resource stretched organisations to share their data (both content and understanding) with the research community, subject to appropriate information governance.

The Data Appliance's will be plug and play into any organisation. They are built upon an open and extendable architecture allowing for innovation and expansion, allowing for full integration with a mix of database systems such as DB2, PostgreSQL, MySQL, Microsoft SQL, and Oracle. The Data Appliance's developed for the initial release are "Dataset Upload and Storage", "Dataset Documentation", "Dataset Descriptive Metrics", "Dataset Validation", "Dataset Quality Reports ", "Dataset Loader and Linkage for SAIL Databank" and "DICOM Image Anonymisation".

The Data Appliance represents a major strategic investment by the SAIL Databank and is shaping the development roadmap. Various key services within the SAIL databank are now delivered using the Data Appliance's.

We will describe the current implementation, the benefits it is delivering, and define the ambitious vision to be delivered under the CIPHER program. We are excited to have formed alliances with key organisations such as Manitoba Centre for Health Policy (Canada) and Curtin University (Australia) enabling use to include the best software and approaches developed within the domain of informatics.

Corresponding author email: simon@chi.swan.ac.uk

Data quality and coverage assessments for the Secure Anonymous Information Linkage databank

Demmler, JC, College of Medicine, Swansea University
Brooks, CJ, College of Medicine, Swansea University
Lyons, RA, College of Medicine, Swansea University

Assessing the quality of data is an imperative and necessary task, especially when working with routinely collected data, however it is not an easy one. Most health records have historically been collected for administrative purposes and were never intended to be used in research. They stand therefore in stark contrast to the definition of high quality data, namely "being 'fit for use' in their intended operational, decision-making and other roles" [2]. Many data analysts will at some point realise, that their data is too messy and do subsequently require major data cleansing [1]. This is not only time intensive, but also requires skilled staff and is subject to skewing of the original dataset (in particular when imputing data).

Some of the problems experienced by data analysts whilst using routinely collected data are as follows:
(i) duplication of records, (ii) missing data, (iii) typing errors, (iv) individual differences in coding practice, (v) working with different GP systems, (vi) spatial and temporal differences in data coverage as well as mismatched of individuals between different datasets.

To be aware of the limitations of the data in the first instance and to ensure that both the data quality and coverage are reasonable for the research project in question, would limit the need of data cleaning and data augmentation. With new data flowing into SAIL over time this is no easy task.

We are presenting a flexible and reproducible methodology to explore the Secure Anonymous Information Linkage (SAIL) databank at the College of Medicine, Swansea University through highly customisable and increasingly automated summary reports. This can be achieved through embedding code of the R Statistical Environment within the LYX Document Processor (a process called sweave). We then use an ODBC connection to connect to SAIL and to import the data of interest through SQL coding into R. This enables us to visualize our data, taking into account changes in time and space. The output report can then be saved as PDF or HTML file. Once implemented as a standard methodology within our research group this methodology will allow us to seemingless include part or all of this exploration in official reports, publications and the SAIL Online Data Dictionary.

References

[1] T.N. Herzog, F.J. Scheuren, and W.E. Winkler. Data quality and record linkage techniques. Springer London, Limited, 2007.
[2] J.M. Juran and A.B. Godfrey. Juran's quality handbook. Juran's quality handbook, 5e. McGraw Hill, 1999.

Corresponding author email: j.demmler@swansea.ac.uk

The care.data programme: developing a modern data linkage service for health and social care in England

Lewis, GH, NHS England, Patient and Information Directorate
Oppenheim, DE, NHS England, Patient and Information Directorate
Hannah, X, NHS England, Patient and Information Directorate
Townend, IJ, NHS England, Patient and Information Directorate
Hallam, V, NHS England, Patient and Information Directorate
Moore, D, NHS England, Patient and Information Directorate
Thomson, K, NHS England, Patient and Information Directorate
Farndon, K, NHS England, Patient and Information Directorate
Flynn, P, NHS England, Patient and Information Directorate
Kelsey, T, NHS England, Patient and Information Directorate

In England, Hospital Episode Statistics (HES) and associated datasets are national resources that have led to tens of thousands of audits and peer-reviewed articles. Yet, two major shortcomings limit their utility: HES data are restricted to hospital care; and the hospital data themselves are incomplete.

NHS England is launching an ambitious plan to create a modern data service for the health and social care system. Known as care.data, this programme will transform HES into a new Care Episodes Service (CES). CES will deliver linked data from primary care, hospitals, community, social care, and mental health settings under the highest standards of quality assurance and information governance.

In May 2013, NHS England published the technical specification of the GP data component of CES. Then, in July 2013, we launched an open consultation on enriching the data extracted from hospitals. In the coming years we will consult on the incorporation of data from additional sources.

In this presentation we will consider (1) the background and goals of the care.data programme; (2) incoming data sources, standards, and time frames; and (3) the outputs, which will include aggregated data for publication, pseudonymised data for limited access, and personal confidential data for patients to download.

Corresponding author email: david.oppenheim@nhs.net

Assessing and documenting bias and error in the linkage of Avon Longitudinal Study of Parents and Children participants to their secondary care records.

Boyd, A, University of Bristol

Longitudinal studies are making increased use of routine health and administrative data as a means of informing missing data techniques and sustainable data collection. These advantages are dependent on the accurate interpretation of the linkage. Links between an individual and their routine records are established by comparing personal identifiers common to both datasets. The potential to do this accurately is impacted by the choice and application of the linkage algorithms and the quality and discriminatory potential of the available identifiers. Recent work by Goldstein, Harron and Wade (2012) demonstrated new methods to enhance the efficiency of the linkage process using multiple imputation (MI) techniques. Once linked, the onus is on the study team to provide the provenance of the data; describing the linkage methodology and assessing the quality of the linkage at an individual level. This paper uses a work-in-progress to link participants of the Avon Longitudinal Study of Parents and Children (ALSPAC) to their secondary health care records as an exemplar of these issues.

ALSPAC is a longitudinal birth cohort which enrolled over 14,500 pregnant women resident in and around the City of Bristol (South West UK) and due to deliver between 1st April 1991 and 31st December 1992. Through the Project to Enhance ALSPAC through Record Linkage (PEARL) we are linking the study index children to their secondary health care records, held within the Hospital Episodes Statistics (HES) dataset. The accuracy of this linkage is of concern as the personal identifiers held in early HES data (pre 1997) will in some cases lack the discriminatory power to identify a single individual. The NHS Data Linkage Service linked a pilot sample of 3,198 study participants to their 1991-2012 HES records. The linkage algorithm varied depending on the ability of the identifiers to establish a “true match”. Where the match was unequivocal a single individual to record link was established. Where there was doubt about the match the study participant was linked to all possible matching records (above a certain match threshold); the information from all these records are made available to the researcher. Using self-reported and administrative study data we will attempt to distinguish between failure to match due to no record existing and failure to match due to error and imprecision in the linkage identifiers. In the case of the later, standard MI methods will be used to impute an outcome. The challenges arising from this linkage, the methodologies used and preliminary findings are discussed.

Corresponding author email: a.w.boyd@bristol.ac.uk

Evaluating bias due to linkage error in anonymised linked data

Harron, K, Institute of Child Health, UCL
Gilbert, R, Institute of Child Health, UCL
Muller-Pebody, B, Public Health England
Goldstein, H, Institute of Child Health, UCL
Wade, A, Institute of Child Health, UCL

Background Errors that occur due to uncertainty in linkage of incomplete or imperfect identifiers can have serious effects on analyses of linked data. Examples in the literature include severely underestimated rates, biased mortality ratios due to differential linkage by ethnic group, exclusion of vulnerable populations due to poorly-recorded identifiers and erroneous rankings of relative hospital performance due to differing data quality between units. The requirement to separate linkage and analysis of linked data to protect patient confidentiality can make it difficult to evaluate bias due to linkage error when reporting results.

Methods We used all available identifiers to link two national databases: PICANet and Labase2. We explored four methods for identifying bias due to linkage error determined what information should be provided by data linkers so that analysts can take any biases into account. The methods comprised: i) Sensitivity analyses using different probabilistic thresholds to provide a range of plausible results; ii) Comparisons of linked and unlinked data characteristics to identify potential sources of bias; iii) Prior-informed imputation, incorporating information from all candidate records, to statistically handle uncertainty associated with choosing only the highest weighted record pairs; iv) Estimates of bias in results, based on comparisons with a subset of gold-standard data.

Results Making available, in an anonymised form, all candidate records (linked and unlinked) with match weights or probabilities would allow i) sensitivity analyses ii) comparisons of data characteristics and iii) prior-informed imputation. Providing the true-match status of a subset of records, obtained from a subset of gold-standard data, would allow analysts to estimate and adjust for bias due to linkage error.

Conclusions Evaluation of linkage performance is vital for the identification of potential sources of bias in linked data. Whilst the physical separation of identifiers and clinical data may be important for privacy protection, the iterative process of estimating match weights or probabilities, evaluating linkage error and adjusting linkage criteria according to the required analysis means that linkage and analysis should be performed in close collaboration. Data providers and analysts need to develop linkage plans that allow metrics of linkage uncertainty to be transferred alongside anonymised data. Meaningful linkage evaluation can improve the reliability and transparency of results based on linked data.

Corresponding author email: katie.harron.10@ucl.ac.uk

Matching using anonymised data

Jones, P, Office for National Statistics

The Office for National Statistics is currently taking a fresh look at options for the production of population and small area socio-demographic statistics for England and Wales. The Beyond 2011 Programme has been established to carry out research on the options and to recommend the best way forward to meet future user needs.

Beyond 2011 is considering a range of options including census, survey and administrative data solutions. Since 'census-type' solutions are relatively understood most of the research is focussing on how surveys can be supplemented by better re-use of 'administrative' data already collected from the public.

The administrative data options being progressed in Beyond 2011 include large scale record linkage between administrative sources and surveys. There is a requirement to optimise the quality of record-linkage under these options as higher match rates are likely to improve the estimation process in the production of population estimates.

It is recognised that the planned approach of matching multiple administrative sources in Beyond 2011 will elevate the associated risks relating to the privacy of data about people and households. It has therefore been a requirement of the programme to consider the development of procedures that will preserve the privacy of individuals’ administrative records.

This paper/presentation will set out the matching strategy that has been developed to address this challenge along with a summary of research undertaken to ascertain the level of quality loss that is incurred by anonymising data prior to record linkage.

Corresponding author email: peter.jones@ons.gsi.gov.uk

Record Linkage Approach in the Dutch Biolink Project

Ariel, A, GGZ inGeest and Department of Psychiatry , VU University Medical Center, Amsterdam, The Netherlands
Bakker, BFM, Statistics Netherlands, The Hague, The Netherlands
Grootheest, G, GGZ inGeest and Department of Psychiatry , VU University Medical Center, Amsterdam, The Netherlands
Laan, DJ, Statistics Netherlands, The Hague, The Netherlands
Smit, JH, GGZ inGeest and Department of Psychiatry , VU University Medical Center, Amsterdam, The Netherlands
Verkerk, ECM, GGZ inGeest and Department of Psychiatry , VU University Medical Center, Amsterdam, The Netherlands

The Dutch Biolink Project aims to improve efficiency by providing a national infrastructure for all medical and socioeconomic record linkages in the Netherlands. We face two major challenges on linking the biobanks data to medical and socioeconomic registries. First, the statutory legal framework requires legal permission from research subjects through an informed consent and prohibits, for instance, the use of sensitive identifiers without encryption. Second, the linkage key has to be constructed from personal identifiers as no common identifier exists throughout the registration systems. At this moment, different organizations apply different linkage keys depending on how they perceive the data quality at hand. The more variable used, the better. However, such approach will make linkage with new data sources rather difficult to realize especially when the variable quality varies.

In this paper we examine which combinations of personal identifiers are indispensable to obtain an acceptable degree of correct links, given the variations in the population covered by registries, the errors present in the identifiers, and the privacy considerations. We examine, among others, whether or not identifier surname should be used in combination with other identifiers, as surname is highly sensitive and possibly not allowed for use, even when it is encrypted. This identifier is known to be discriminative yet more susceptible to error, i.e., prone to both misspelling and swapping between first and last name. Because the actual linkage will be performed based on encrypted values, partial agreement between identifiers will also be used to reduce the influence of errors. To link the records, we apply both deterministic and probabilistic methods. We employ a mathematical programming approach in the probabilistic method in order to maximize the number of correct links.

For testing purposes, due to data privacy concerns, we will first test the linkage using simulation datasets. In collaboration with the organizations involved, we develop simulation datasets based on their data. Each of these dataset represents a certain population subgroup. We also add typical errors seen in real data at varying degrees to study their effect on the linkage methods. The number of false negatives and false positives will be used to measure the linkage keys performance, as well as the extent to which the probabilistic method outperforms the deterministic method.

Corresponding author email: a.ariel@ggzingeest.nl

Recording of acute myocardial infarction events in primary care, hospital admission, disease registry, and national mortality records

Herrett, E, London School of Hygiene and Tropical Medicine
Shah, AD, UCL
Boggon, R, Clinical Practice Research Datalink
Denaxas, S, UCL
Smeeth, L, London School of Hygiene and Tropical Medicine
van Staa, TP, London School of Hygiene and Tropical Medicine
Timmis, A, Barts and the London School of Medicine and Dentistry
Hemingway, H, UCL

Objective: To compare recording, risk factors, mortality and diagnostic validity of acute myocardial infarction in primary care, hospital discharge, disease registry and mortality records.

Design and participants: a sample of patients with acute myocardial infarction in England between January 2003 and March 2009, identified in four prospectively collected, linked electronic health record sources: the primary care Clinical Practice Research Datalink, Hospital Episode Statistics, disease registry (Myocardial Ischaemia National Audit Project) and Office for National Statistics cause-specific mortality data.

Setting: One country (England) with one health system (the NHS).
Main outcome measures: recording of acute myocardial infarction, incidence, all-cause mortality within one year of acute myocardial infarction, diagnostic validity of acute myocardial infarction compared to electrocardiographic and troponin findings in the disease registry (gold standard).

Results: We found 21 482 patients with a record of acute myocardial infarction in one or more sources. Risk factors and non-cardiovascular co-existing conditions were similar across patients identified in primary care, hospital admission and registry sources. Immediate all-cause mortality was highest among acute myocardial infarction patients recorded in primary care, which (unlike hospital admission and disease registry) includes patients who did not reach hospital, but at one year mortality rates in cohorts from each source were similar. 5561 (31.0%) of patients with non-fatal acute myocardial infarction were recorded in all three sources and 11 482 (63.9%) in at least two sources. Crude incidence of acute myocardial infarction was underestimated by 25-50% using one source compared to using all three. Compared to acute myocardial infarction defined in the disease registry, the positive predictive value of acute myocardial infarction recorded in primary care was 92.2% (95% CI 91.6, 92.8) and in hospital admissions was 91.5% (95% CI 90.8, 92.1).

Conclusion: Failure to use linked electronic health records from primary care, hospital, disease registry and death certificates may lead to biased estimates of myocardial infarction incidence and outcome.
Cohort study registration: This CALIBER study is registered on clinicaltrials.gov (unique identifier NCT01569139).

Corresponding author email: emily.herrett@lshtm.ac.uk

The CALIBER Platform: phenotyping raw linked Electronic Health Record (EHR) data at scale for translational research

Denaxas, S, Clinical Epidemiology Group, Department of Epidemiology and Public Health, UCL
Pujades, M, Clinical Epidemiology Group, Department of Epidemiology and Public Health, UCL
Shah, A, Clinical Epidemiology Group, Department of Epidemiology and Public Health, UCL
Kalra, D, CHIME, Institute of Epidemiology & Health
Hemingway, H, Clinical Epidemiology Group, Department of Epidemiology and Public Health, UCL

Background

The Cardiovascular Disease Research using Linked Bespoke Studies and Electronic Health Records (CALIBER) programme [1] is part of the London Health eResearch Centre (HeRC), CHAPTER. CALIBER currently links electronic health record (EHR) data from the Clinical Practice Research Datalink, Hospital Episode Statistics, the Myocardial Ischaemia National Audit Project (MINAP) and Office of National Statistics mortality data in >2m patients.

Raw linked EHR data suffer from variable data quality, with data being missing, erroneous or repeated across sources. Each EHR source makes use of a different coding system (Read, ICD9/10, OPCS-4), resulting in data of different granularity that are hard to pool, query and analyse at scale. These challenges limit the large-scale automated use of linked EHR data for research.

Processing is required to ensure that raw data are transformed into research-ready variables that are an accurate representation of the clinical data and make use of all available sources. Metadata should be made available so variables are reproducible by researchers. Similar approaches are used by, for example, the EMERGE Network but without population-wide, or primary care EHR data.

Objective

To create a platform for deriving and validating clinical phenotypes from raw EHR data for use in translational research at scale.

Methods

We have developed a data-driven approach for phenotyping raw EHR data and transforming them into research-ready variables of clinically relevant features. Variables make use of all EHR sources, and definitions are agreed by clinical and non-clinical researchers. Generated metadata are published ensuring transparency and enabling researchers to reproduce them.

Results

We have created >470 variables on medical history, diagnosis, investigations, procedures and prescriptions which are catalogued by ICD10 chapter heading.

For each variable the Data Portal provides: the clinical codes defining it, programming scripts used to extract the data, variable metadata, implementation details, references to published literature using the variable and other notes. An example of such a variable is the Atrial Fibrillation CALIBER phenotype (see abstract by Wallace J).

Discussion

The Data Portal is available at http://www.caliberresearch.org and contains data, metadata and tools on over 470 variables. It is currently being used across partner institutions in the HeRC, including UCL and LSHTM. All data on the Portal are made available for free under a Creative Commons Attribution licence subject to user registration. Users are encouraged to contribute new algorithms to enhance the range of research variables available.

References: *[1]Denaxas S., Int. J. Epid. 2012;41:1625-1638.*

Corresponding author email: s.denaxas@ucl.ac.uk

Automated phenotyping using free text in primary care electronic health records in patients with coronary artery disease: a CALIBER study

Shah, AD, Clinical Epidemiology Group, Department of Epidemiology and Public Health, UCL, London, UK
Denaxas, S, Clinical Epidemiology Group, Department of Epidemiology and Public Health, UCL, London, UK
Pujades, M, Clinical Epidemiology Group, Department of Epidemiology and Public Health, UCL, London, UK
Riedel, S, Department of Computer Science, UCL, London, UK
Hemingway, H, Clinical Epidemiology Group, Department of Epidemiology and Public Health, UCL, London, UK

BACKGROUND

Primary care electronic health records are a rich data source but much information is in free text rather than in coded form. Deeper phenotyping in cohort studies (e.g. UK Biobank) or trial follow-up using primary care records may require the use of free text information. For example, the type of myocardial infarction (ST elevation, STEMI or non-ST elevation, NSTEMI) and degree of left ventricular impairment are rarely available in the coded data.

METHODS

We piloted a method of extracting this information by modifying the Freetext Matching Algorithm (FMA), a natural language processing system for clinical text developed by our group that maps text phrases to Read codes using tables of synonyms and context patterns [1]. We used a sample of 1872 anonymised free text records from a previous coronary artery disease study using the Clinical Practice Research Datalink (CPRD). We reviewed a random subset of 554 texts and added entries to FMA lookup tables to enable it to classify type of myocardial infarction and left ventricular function, and tested the modified FMA on the remaining 1318 texts. We compared recall using FMA versus the existing Read codes associated with the texts, with manual review by a clinician as the gold standard.

RESULTS

Only four records had a specific Read code for type of myocardial infarction, but FMA detected 10/11 NSTEMI and 9/12 STEMI diagnoses, with precision (positive predictive value) 100% and recall (sensitivity) 83% (95% CI 61%, 95%). FMA tended to miss diagnoses expressed in a non-standard way such as "st elev. Inferior leads, depression v2 - 6. Acute infarct". Eighty-three records contained information on left ventricular function, but only two had specific Read codes. FMA achieved 97.5% precision and 65% recall (39/60; 95% CI 52%, 77%) for detection of any grade of left ventricular dysfunction, and 95.5% precision and 49% recall (21/43; 95% CI 33%, 65%) for classifying the grade (mild, moderate or severe).

DISCUSSION

FMA had high precision but moderate recall for detecting left ventricular function or type of myocardial infarction in free text. Although FMA has difficulty interpreting complex language structures, our findings suggest it is a useful method which can be adapted for other clinical areas. We plan to scale up this approach for patients with myocardial infarction in the CALIBER programme [2].

REFERENCES

1. Shah AD, Martinez C, Hemingway H. BMC Med Inform Decis Mak. 2012;12:88.

2. Denaxas SC, George J, Herrett E, et al. Int J Epidemiol. 2012;41:1625-1638.

Corresponding author email: a.shah@ucl.ac.uk

Outcomes following Primary knee replacement

Lyons, R.A, Centre for Improvement of Population Health through E-records Research (CIPHER), Swansea University, Swansea, UK
Palmer, S.R, Institute of Primary Care and Public Health, Cardiff University, Cardiff, UK
Griffiths, S, Cardiff and Vale University Health Board/Public Health Wales
Walters, A.M, Centre for Improvement of Population Health through E-records Research (CIPHER), Swansea University, Swansea, UK

Introduction

The Secure Anonymised Information Linkage (SAIL) databank has been used to map orthopaedic care pathways across primary and secondary care. The work has focussed on orthopaedic care pathways for knee pain. The project involved identifying individuals with first time knee pain, and finding the proportion of those who attended Trauma and Orthopaedic Outpatients over time and the proportion who received a primary knee replacement over time. Outcomes following primary knee replacement also form part of the care pathway which this presentation will focus on. The aim is to provide outcome measures that will inform clinicians and that can be developed to be used as a tool for shared decision making with patients. The outcomes analysed include revision procedures, complications and mortality.

Method

Using the Patient Episode Database for Wales (PEDW) data held in the SAIL databank, a cohort of individuals who received their first primary knee replacement between 01/01/2006 and 31/12/2010 were identified. Those who received a primary knee replacement were linked back with the PEDW data to identify the individuals who developed a complication and those who received an elective knee revision. The Welsh Demographic Service dataset was used to obtain the date of death. Survival analysis was carried out to find the risk of developing a complication, the risk of receiving a revision and the risk of mortality within a certain time period following a primary knee replacement.

Results

Preliminary results show that the estimate of the risk of developing a complication 4 years following surgery is approximately 10%. The estimate of the risk of receiving a revision 4 years following surgery is just over 3%. Those aged less than 65 have a higher risk of receiving a revision. The estimate of the risk of death 4 years following a primary knee replacement is approximately 7.5%. Males and those aged >=75 are greatest at risk.

Corresponding author email: a.m.walters@swansea.ac.uk

Social inequality and hospitalisation costs for the first year of life of preterm infants

Cannon, JW, Telethon Institute for Child Health Research
Langridge, A, Telethon Institute for Child Health Research
Glauert, R, Telethon Institute for Child Health Research

The cost of health care is continuously increasing as a proportion of GDP. In the 2008/09 financial year, the Australian national average public hospitalisation cost for neonates and other newborns increased by 7.3% compared to an average of 5.8% for all hospitalisations. A large proportion of these costs are due to preterm birth (< 37 weeks gestation), which can be influenced by lifestyle factors and social inequalities. This suggests that costs could be reduced by effective health promotion strategies, as well as policies aimed at reducing social inequalities. This study utilised population-level data from the Western Australian Data Linkage System to quantify the hospitalisation costs within the first year of life for infants born preterm in Western Australia (WA).

Obstetric and birth data were extracted from the WA Midwives Notification System and linked to hospitalisations and geographical location recorded in the WA Hospital Morbidity Data System for infants born in the financial years 2006/07 and 2007/08. Social inequality was measured by the Australian Socioeconomic Indexes for Areas using the Index of Education and Occupation (IEO), with a higher score indicating a higher level of education and occupation. Hospitalisation costs were assigned according to Australian Refined Diagnostic Related Group cost weight reports for WA, adjusted to constant 2012 prices and aggregated for each separation. Differences in the total mean cost of hospitalisations between socio-economic groups were tested for statistical significance by ANOVA techniques. The sequence of hospitalisations within the first year of life was categorised into none, neonatal admission(s) only (<=28 days since birth), infant admission(s) only (>28 days since birth) and neonatal with further ongoing admissions, with differences between socio-economic groups assessed using the chi-square statistic.

Preliminary analysis indicated an inverse relationship between mean total cost of admissions within the first year of life and IEO level, except for infants born to parents from the top IEO decile which had elevated costs similar to the population mean. Infants born preterm to parents from the top IEO decile were more likely to require only an initial neonatal admission without future hospitalisations while those born to parents in the bottom IEO decile were more likely to require initial neonatal hospitalisations with ongoing hospitalisation.

The reasons for these cost differences between IEO levels need to be further investigated, however, these results suggest a cost-effective analysis looking at potential health promotion and education strategies which targets both parents of the lowest and highest socio-economic.

Corresponding author email: jcannon@ichr.uwa.edu.au

Not all high users are created equal: Correlates of higher-than-expected health care services use in British Columbia, Canada

McGrail, KM, Centre for Health Services and Policy Research, The University of British Columbia

A small proportion of users accounts for a large share of health care expenditures. High users tend to have complex conditions but there are also wide variations in care. This analysis brings together the “high user” and “variations” literature by assessing correlates of higher than expected service use.

This analysis used linked, individual-level data from five distinct sources: physician payments, hospital separations, pharmaceutical prescriptions, a patient demographics file, and a physician demographics file, all from 2009/10. Physician and pharmaceutical costs were available in the data sets. Hospital costs were estimated using the Canadian Institute for Health Information’s resource intensity weights. Observed expenditures for each individual were calculated overall and for 8 sub-components (GPs, medical specialists, surgical specialists, laboratory, imaging, acute care, day surgery and pharmaceuticals). Expected expenditures overall and for each sub-component were estimated with a two-part approach implemented using the “nlmixed” procedure in SAS. Modeling variables included age, sex and ACG, a validated estimate of need for health care services. We used regression to identify patient and physician characteristics associated with having higher than expected expenditures.

As expected, health care expenditures are highly skewed. The top 5% of users (n~200,000) accounted for more than 50% of total expenditures (>$4billion). Costs per capita for high users are remarkably similar across age groups. What differs is the proportion of individuals within an age group who are in the overall top 5%, ranging from about 1% for young children and adolescents to nearly 30% of those aged 85+. There is considerable variation in the difference between observed and expected values (put differently, there is variation in health system costs not explained by age, sex, and health status). Variables reflecting health system (having a regular family doctor), socioeconomic (income decile), and geographic (rurality) characteristics all contributed to explaining the remaining variation.

There are differences in health care expenditures cannot be explained by basic demographics or need for health care services. Some of these differences appear to be systematic in the sense that they relate to identifiable characteristics of individuals (e.g. location of residence) or the system (e.g. having a regular doctor) that a priori would not be hypothesized to influence need. Further research will assess the extent to which these differences are mediated by individual and/or regional physician practice styles.

Corresponding author email: kmcgrail@chspr.ubc.ca

Optimizing the identification of related episodes in routinely collected, non-personalized inpatient data based on record linkage

Endel, F, Technical University of Vienna, Austria
Katschnig, H, Ludwig Boltzmann Institut for Social Psychiatry, Vienna, Austria

Routinely collected data (e.g. reimbursement data) about hospital episodes is often used as a basis for national and international statistics on health service use. These statistics (as published by e.g. OECD, Eurostat) are usually episode and institution, and not patient centered. To circumvent the shortcomings of these statistics, record linkage approaches are used to generate patient centered information and to cover other parts of the health care system like the outpatient sector as well. In comparison to anonymised data from the inpatient sector, personalized or at least pseudonymized datasets are often harder to acquire or not available at all. Additionally concerns about data privacy and possible changes of the legal circumstances in the EU exist.

Therefore record linkage methods which allow the identification of related episodes in series of routinely collected and non-personalized inpatient data are required for enhancing the value of gathered statistics. While it is known how to deal with data quality issues in general, it is hard to get an idea of the proportion of wrongly associated episodes, the importance of different variables and the influence of varying levels of detail.

The presented work is based on the GAP-DRG database, including linked records from all 19 Austrian social insurance institution (on utilization of all services except hospital services) and the non-personalized records of the hospital DRG data base of the Ministry of Health. Based on the knowledge about the correct mapping between inpatient episodes and the availability of detailed personal data from the outpatient sector, the process of finding associations between non-personalized episodes of the DRG system can be analyzed and optimized regarding the needed variables, data quality, different thresholds and the usage of expert knowledge.

As a result it is not only possible to gather pathways of single patients at least in the DRG system over longer time periods, but also to get an idea about common errors, artifacts and required data to perform this task. Additionally this approach prepares for potential future difficulties caused by new privacy regulations.

Corresponding author email: florian@endel.at

Development of an algorithm to identify and describe retroperitoneal lymph node dissections in hospital episode statistics

Evison, F, Health Informatics Department, University Hospitals Birmingham NHS Foundation Trust, United Kingdom
Begaj, I, Health Informatics Department, University Hospitals Birmingham NHS Foundation Trust, United Kingdom
Mak, D, Department of Urology, University Hospitals Birmingham NHS Foundation Trust, United Kingdom
Bolton, R, Department of Urology, University Hospitals Birmingham NHS Foundation Trust, United Kingdom
Viney, R, Department of Urology, University Hospitals Birmingham NHS Foundation Trust, United Kingdom
Patel, P, Department of Urology, University Hospitals Birmingham NHS Foundation Trust, United Kingdom

A retro-peritoneal lymph node dissection (RPLND) is a complex procedure that can form part of a testicular cancer patient’s treatment. The European Association of Urology recommends that an RPLND is performed on post-chemotherapy patients with a non-seminoma germ cell tumour (NSGCT) with residual mass whose serum tumour markers are normal or plateauing.

From patient notes, it was possible to identify 118 patients who had undergone an RPLND at University Hospitals Birmingham (UHB). Unfortunately, coding issues meant that there were multiple ways in which an RPLND was coded on the electronic patient systems. These codes are used to identify procedures in hospital episode statistics (HES) returns. We randomly split the patient data into two based on the month and year of the operation, thus allowing the local HES returns to be divided in the same manor. Using one half of the dataset we produced an algorithm to identify the patients who had had an RPLND. This algorithm was then tested on the other half, and used to calculate sensitivity (90.3%), specificity (99.99%), positive predictive value (91.8%) and negative predictive value (99.99%).

The algorithm was then implemented on the national HES dataset to identify patients who had had an RPLND. We included patients who were admitted between April 2001 and March 2012.We excluded identified patients when we were not sure whether they had had an RPLND due to testicular or penile cancer. We linked the records of the identified patients to ONS mortality records to describe long term survival. We describe the characteristics of patients undergoing this procedure, procedures occurring concurrently to the RPLND, and complications arising from the RPLND. We also linked the data to the HES outpatient dataset to identify those patients receiving chemotherapy as an outpatient.

We identified 1,237 men who underwent at least one RPLND during the period and fewer than seventy of these had a further RPLND. About half of the men had a form of orchidectomy before their RPLND and a tenth had an orchidectomy performed on the same day as the RPLND

The British Association of Urological Surgeons completed a voluntary audit of RPLNDs between April 2012 and March 2013. It is envisioned that we will compare the results obtained using the algorithm on the HES data to those obtained through the audit.

Corresponding author email: felicity.evison@uhb.nhs.uk

Methods for identifying procedural complications from national routine healthcare databases: a literature review

Keltie, K, Newcastle Upon Tyne Hospitals NHS Foundation Trust
Cole, H, Newcastle Upon Tyne Hospitals NHS Foundation Trust
Arber, M, York Health Economics Consortium
Patrick, H, National Institute for Health and Care Excellence
Sims, AJ, Newcastle upon Tyne Hospitals NHS Foundation Trust

Objective: To systematically search for methodologies to identify procedural and device-related complications from national routinely collected data.

Method: The Cochrane Methods Register, Econlit, EMBASE, Health Management Information Consortium, Health Technology Assessment, MathSciNet, MEDLINE, MEDLINE in-process, OAISTER, OpenGrey, Science Citation Index Expanded and ScienceDirect databases were searched.
MEDLINE searches combined free text terms for datasets (Hospital Episode Statistics (HES), General Practice Research Database (GPRD), Clinical Practice Research Database (CPRD), Datix, National Reporting and Learning System (NRLS), National Patient Safety Agency (NPSA), Medicines and Healthcare products Regulatory Agency (MHRA), Office of National Statistics (ONS)) with free text and index terms related to complications (e.g. adverse events, incidents, readmissions, re-attendances, reoperations). Included studies were published since 1987 and described replicable methodologies to identify procedural or device-related complications from routine data. Studies which described drug-related complications, disease incidence, or had unspecified methodology, were excluded.

Results: In total, 3688 abstracts (6049 before de-duplication) were assessed for relevance, independently by two authors with arbitration by a third, leaving 512 of potential relevance for critical review of full text. Of these, 110 presented a replicable methodology and were analysed further. Of the 67 studies using HES, three techniques were reported: 1) 65 searched for named complications from a priori lists using International Classification of Disease (ICD) diagnosis and/or Office of Population Censuses and Surveys Classification (OPCS) codes, 2) 1 extracted diagnosis and procedure codes over an interval and manually identified complications, 3) 1 searched for complications using “Y-” and “T-” ICD complication codes recommended by the NHS Classification service. In the 27 studies using ONS, deaths were identified through data linkage via ONS, or third party organisation. Of the 25 methodologies using the NRLS database, 9 used manual searches of the incident speciality field (i.e. hospital department), 4 used the free-text field, 6 used a combination of both fields, 4 had the search conducted by NPSA and 2 were automated free-text searches using keyword and letter sequences. All 7 studies using GPRD searched for named diagnoses or procedures using Read/OXMIS coding. Two studies used incidents reported to MHRA; one was identified through surveillance of medical device bulletins, the other contacted MHRA directly.

Conclusions: Several different approaches have been used to identify procedural complications from routine data, with HES being the most frequently used. We are currently assessing the advantages and disadvantages of each by applying to specific procedures.

Corresponding author email: kim.keltie@nuth.nhs.uk

Combining diagnosis, procedure, and drug information from primary and secondary care to define clinical phenotypes: a case study of Atrial Fibrillation in the CALIBER programme

Wallace, JS, PhD Student on MRC PROGRESS Partnership; Clinical Epidemiology Group, Department of Epidemiology and Public Health, UCL
Morley, KI, Clinical Epidemiology Group, Department of Epidemiology and Public Health, UCL
Patel, R, Clinical Epidemiology Group, Department of Epidemiology and Public Health, UCL
Denaxas, S, Clinical Epidemiology Group, Department of Epidemiology and Public Health, UCL
Shah, A, Clinical Epidemiology Group, Department of Epidemiology and Public Health, UCL
Perel, P, London School of Hygiene and Tropical Medicine, UCL
Hunter, R, St Bartholomews & The Royal London Hospital, London, UK
Schilling, R, St Bartholomews & The Royal London Hospital, London, UK
Timmis, AD, Barts and the London School of Medicine and Dentistry, London, UK
Hemingway, H, Clinical Epidemiology Group, Department of Epidemiology and Public Health, UCL

Background: Integrating linked primary and secondary care Electronic Health Records (EHRs) in a clinically and analytically valid manner is a major challenge. Atrial fibrillation (AF), a common condition associated with stroke, illustrates this well because diagnosis, drug and procedure information may be available in multiple sources, but previous EHR studies have not addressed this. The majority have focused solely on diagnostic codes, often only examining hospital records and death certificates.

Objective: To develop and validate an algorithm for AF that combines diagnostic and treatment information from primary and secondary care.

Methods: Four cardiologists were consulted to identify diagnostic and treatment codes for AF in primary and secondary care EHRs. From these we developed a multi-stage algorithm to define two categories of diagnosis - a coded diagnosis, and an inferred diagnosis based on the pattern of prescriptions after excluding for alternative indications for use.
We evaluated face validity of the algorithm using data from the CALIBER programme, which links CPRD data from primary care (Read codes) and secondary care (ICD-10 codes for diagnoses, OPCS procedure codes) [Denaxas et al. (2012) Int J Epi 41:1625]. We defined a cohort of adults without AF aged ⩾30, and examined the association of known risk factors at baseline with incident AF. Further examination of these associations using Cox proportional hazards models is ongoing.

Results: There is one ICD code (I48) for AF and 21 Read codes for current or historical AF; these defined our coded diagnosis category. Our inferred category was defined by prescriptions for warfarin (in the absence of pulmonary embolism or deep vein thrombosis) or digoxin (in the absence of heart failure).
Applying our algorithm to the CALIBER cohort of 2,086,066 patients (median follow-up 6 years) identified 72,885 coded diagnoses of AF from primary care (32%), secondary care (51%), and both sources (17%). An additional 8,746 cases were inferred from prescriptions (warfarin 63%, digoxin 37%). Using only secondary care data would have reduced the identified cases by 39%.
We confirmed strong associations with the known AF risk factors of heart failure, hypertension and obesity. Age and sex adjusted ORs for diagnosed AF were 3.4, 1.5 and 1.7 respectively, and 3.6, 1.5, and 1.5 for inferred cases.

Discussion: For many conditions like AF, patients may neither be hospitalised nor have a recorded diagnosis; focus on a single EHR source will result in reduced case identification. Our implementation of an algorithm for identifying AF cases that integrates data from primary and secondary care illustrates the value of the approach taken in CALIBER.

Corresponding author email: joshua.wallace.12@ucl.ac.uk

Overcoming methodological challenges in estimating inpatient exposure to nurse staffing by linking nursing payroll data to hospitalisation records

Schreuders, L W, School of Population Health, The University of Western Australia
Bremner, A, School of Population Health, The University of Western Australia
Geelhoed, E, School of Population Health, The University of Western Australia
Finn, J, Faculty of Health Sciences, Curtin University

There is ongoing professional, academic and political interest in research quantifying the unique contribution of nursing to health outcomes. Research on this topic in acute hospital settings has found an inverse association between nurse staffing and certain inpatient complication rates, known as nursing sensitive outcomes (NSOs). Inpatient complications are said to be NSOs if their prevention falls within the scope of usual nursing practice. For example, a patient's risk of developing a pressure ulcer (bedsore) increases if nurses do not regularly reposition patients and perform skin integrity assessments.

This project aimed to use data linkage to address two methodological challenges frequently cited in the literature, accurate ascertainment of both nurse staffing and NSO incidence rates. Four administrative data sources were linked; mortality records, inpatient hospitalisation records, inpatient ward movements, and nursing payroll data.

The first challenge, accurate measurement of nurse staffing, has been addressed by measuring staffing per nursing shift (three times per day) but this requires prohibitively resource intensive data collection. Researchers have also estimated hospital-level nurse staffing averaged over months or years. Inability to capture the varied ward-level staffing within each hospital is one of several limitations with this method (i.e. because patient care requirements on an intensive care unit necessitate higher staffing than on a rehabilitation ward). Linkage of nursing payroll to patient ward census data is a novel approach which utilises existing administrative data and results in detailed measurement of nurse staffing per hospital ward.

The second challenge is detecting when NSOs have occurred using administrative data sources. Algorithms currently used to distinguish whether inpatient complications are the result of the underlying health status of the patient (rather than the nursing care received) treat each hospitalisation as an independent event. Using linked data to include information from earlier hospitalisations allows for a better understanding of each patient's underlying health status. The results of this project suggest that if algorithms that detect NSOs do not include data from earlier hospitalisations, inpatient complications are over-attributed to nursing care.

This presentation will describe linkage to a data source not routinely accessed for health research, a method which could be replicated and refined in varied patient care settings internationally. The lessons learned in linking the four data sources to determine patient-level exposure to nurse staffing and NSO incidence will be presented.

Corresponding author email: louise.schreuders@uwa.edu.au

Developmental Pathways Project: Challenges for Western Australia Data Linkage

Quintero Silvestre, M, Western Australia Department of Health and Telethon Institute for Child Health Research

Data linkage research is a fundamental part of societal development. The changing nature and complexity of research projects consistently observed over the last decade have introduced significant challenges to data linkage models. Balancing communication, confidentiality, privacy and timely data release have become a core issue for researchers, data custodians and the community. This presentation will illustrate some of the complexities of the Developmental Pathways Project (DPP) and will present some of their effects on the Western Australian (WA) data linkage model.

The very unique nature of the research undertaken by DPP has introduced several challenges to the data linkage model in WA. Firstly, the project involves a large collaboration of government agencies and sources data from all of them. Consequently, it has required the creation of a linkage infrastructure which has been maintained since the inception of the project, nine years ago. This infrastructure involves linkage of numerous databases otherwise not accessible for research. The increased number of datasets utilised in a single project and a wider use of population-based cohort designs has significantly broadened the complexity of the research. Such complexity has been associated with added constraints for approving and releasing data due to privacy and confidentiality issues. Furthermore, the process of data extraction prior to its release usually implies duplicate efforts from the Data Collections involved. The combination of the above issues has led to less efficient processes for delivery of data and the timely commencement of studies.

Joint efforts among agencies have been made to deal with the growing complexity of the project and to counteract the implications of more rigorous approval processes. Some of the strategies carried out include the establishment of new communication conduits and development and implementation of data resources for custodians and researchers. These resources include a tool for centralising service data extractions and cohort selection (CARES) and a metadata website containing information about the data sources. In addition to this, a consumer reference group has been gradually set up to involve community members. Forthcoming additions to the already established challenges comprise the expansion of the infrastructure by adding datasets from Police and the Department of Housing. It is also expected that negotiations will commence with other non-health government agencies to include their data into CARES.

The above strategies aim to deal more efficiently with the issues raised by the growing demands of complex research projects. They have also contributed to gradually raising awareness among researchers, custodians and consumers of the impact the evolving nature of data linkage research has on balancing confidentiality, privacy and the release of data. It is expected the new strategies sought and implemented in the near future will contribute to addressing the demands of emerging research intricacies.

Corresponding author email: marcela.quinterosilvestre@health.wa.gov.au

Challenges of linking child survey data to other routinely collected data

Turner, SL, Swansea University
Lyons, RA, Swansea University
Rodgers, SE, Swansea University
Fry, R, Swansea University
MacKay, M, European Child Safety Alliance
Vincenten, J, European Child Safety Alliance
McFarlane, K, Children in Wales

As part of the European Project TACTICS (Tools to Address Childhood Trauma, Injury and Children's Safety) Swansea University are leading the development of a new online “child Safety Survey”, designed to collect information about the types of hazards and safety features children are exposed to in their local environment. The survey will be completed by children aged 8 - 13 years on an annual basis, and will predominantly be carried out in school classrooms. It is currently being piloted in several European countries including Wales, and if the survey proves feasible and acceptable, will be one of the first standardised surveys to collect data on child safety in Europe.

Currently the survey does not collect identifiable data; however, in some countries there will be the opportunity to link survey results to other individual data (e.g. health, educational and environmental data) through existing systems in order to explore whether changes in exposures affect outcomes (e.g. health and educational attainment).

In Wales, we plan to link survey results to other individual data using the Secure Anonymised Information Linkage (SAIL) databank. SAIL is a large scale anonymised databank at Swansea University, which enables individual and household level health, educational and environmental data to be linked together and analysed to support health related research. Linking these survey results to the SAIL database will be a major development, as it will provide much needed evidence on the medium and long-term effectiveness of safety policies and interventions. However, before this is possible a number of challenges need to be addressed.

Although the SAIL “split-file” methodology will be used to ensure the anonymity and confidentiality of children partaking in the survey are maintained; the nature of this work means that ethical approval is a prerequisite. It is also likely that we will require consent from both the school and parents. In addition, younger children completing the survey may struggle to provide accurate personal details; thus, we will require support from the education department/schools, to ensure children are accurately anonymised and incorporated into the SAIL databank.

This conference will provide both a valuable opportunity to discuss the issues faced with this type of data linkage, and a platform to share our approaches with other teams who may be attempting similar work.

Corresponding author email: s.turner@swansea.ac.uk

Electronic Checking Process of Governmental Data: A Qualitative Approach

Pinto, P, PESC/COPPE/UFRJ - Rio de Janeiro - RJ - Brazil
Cerceau, R, National Regulatory Agency for Private Health Insurance and Plans - Brazil
Mesquita, R, PESC/COPPE/UFRJ - Rio de Janeiro - RJ - Brazil
Carvalho, LA, PESC/COPPE/UFRJ - Rio de Janeiro - RJ - Brazil

INTRODUCTION

The Natural Persons Register (NPR) and Beneficiary Register (BR) are both Brazilian nationwide governmental databases maintained by the Federal Revenue of Brazil and the National Regulatory Agency for Private Health Insurance and Plans (NRAPHIP), respectively. The NPR stores personal records of tax payers in Brazil. The BR stores contractual records of beneficiaries (people who can afford some sort of private healthcare services in Brazil).

In 2010, 34 millions of records in the BR matched to a similar one in the NPR. The NRAPHIP took the NPR number (similar to the social security number in the U.S.) as the blocking factor. The equality of a phonetic transcription of the first and the last name; and birth date were the matching criteria of that achievement.

In this work the concept of Electronic Checking Process (ECP) is presented. It is a theoretical guideline to perform an automated comparison between database records and qualitative assignment of categories, which is a conceptual improvement of above mentioned method. From the ECP concept, a record linkage process (RLP) is derived. This one uses only personal identification information (full name, birth date and personal identification number).

METHODOLOGY

Five conditions to implement an ECP are:

1) Coherence principle: Given two distinct records of information, there is a set of common attributes that, in principle, ensures that these two records refer to the same entity. If these two records possess such kind of set in common they are called coherent in principle;

2) Potentiality of Variation: Given two records coherent in principle, there may be a set of attributes that can vary (called analogous attributes). These attributes represent properties, or characteristics, of the referenced entity that can be compared. If these two records possess such kind of set in common they are called potentially variants;

3) Rationalization of Inconsistencies: The existence of a priori knowledge with the potential to provide at least one plausible explanation of why two coherent in principle records are potentially variants;

4) Comparable Qualification: The existence of qualifying function that assigns a comparable degree to a plausible explanation;

5) Automatic Execution: Items 1- 4 have algorithmic nature.

RESULTS

The QUALISDATA RLP makes only few assumptions about the quality of pair of records being matched. The quality of the matching pair is measured by a cognitive response simulated by comparisons, heuristics and metrics. Its qualifying function behaves smoothly and similarly to a credit-rating indicator.

Corresponding author email: pcoelhopinto@cos.ufrj.br

The prevalence of chronic conditions in children who die: estimates based on death certificates linked to longitudinal hospital admission data

Hardelid, P, Centre for Paediatric Epidemiology and Biostatistics, UCL Institute of Child Health
Dattani, N, Centre for Paediatric Epidemiology and Biostatistics, UCL Institute of Child Health
Davey, J, Centre for Paediatric Epidemiology and Biostatistics, UCL Institute of Child Health
Gilbert, R, Centre for Paediatric Epidemiology and Biostatistics, UCL Institute of Child Health
on behalf of the RCPCH Programme Board and The Child Death Overview Working Group, ,

The contribution of chronic health problems to childhood mortality is becoming increasingly important as mortality from injuries and infections in healthy children decline, and survival of babies born with complex problems improves. While immediate causes of mortality are listed on death certificates, these may not capture chronic conditions affecting long-term survival of children.

We determined the prevalence of chronic conditions in children aged one to 18 years who died between 2001 and 2010 using death certificates linked to the child's trajectory of past hospital admission records in England, Scotland and Wales. We developed criteria to identify chronic conditions on hospital and death certification records and classify these into eight groups based on a literature review and iterative review of coding clusters by researchers and a panel of clinicians. We determined the added value of data from the hospitalisation trajectory on the prevalence of chronic conditions in children who die by successively adding more data, starting with the underlying cause on the death certificate and adding hospitalisation data from previous years.

Over three-quarters of children who died could be linked to one or more hospital records (79% of 20104 in England; 92% of 2391 in Scotland; 77% of 943 in Wales). The prevalence of one or more chronic conditions at death was highest in children aged less than 10 years (79.0% of 9328 children aged less than 10 years compared to 70.5% of 14110 children aged 10-18 years), and higher in girls than in boys (79.3% of 9191 girls aged 1-18 years and 70.4% of 14247 boys of the same age) going back to the first available hospital record. Death certificates underestimated the prevalence of most chronic conditions. For example, the proportion of children who died with neurological conditions, the most common of the chronic condition groups, was 25.1% (5880/23438, 95% confidence interval (CI) 24.5%-25.6%) using death certificates only, but increased to 43.2% (10126/23438, 95% CI 42.6%-43.8%) if a child's hospital records were included (going back to their earliest available hospital record).

Linkage to longitudinal hospital admission data is needed to adequately measure chronic underlying conditions in children. These chronic conditions may not directly cause death but may be relevant to their deterioration and may indicate which specialties are responsible for their care.

Corresponding author email: p.hardelid@ucl.ac.uk

Childhood and Early Adulthood Predictors of Mortality: The 6-Day Sample of the Scottish Mental Survey 1947

Calvin, CM, Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh
Deary, IJ, Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh

A recent data linkage was conducted of up-to-date vital status records of a Scottish 1936 birth cohort, to data on individuals' psychological traits, education, health, and socioeconomic status, from an assessment period spanning 1947 to 1963. The members of the "6-Day Sample" were 11 when they first took part in a national initiative by the Scottish Council for Research in Education, which tested the cognitive performance of all age 11 Scottish schoolchildren in 1947 (n = 70,805). This special sub-sample of individuals (n = 1208), selected on the basis of being born on the first day of the even-numbered months of 1936, and, representative of the background population, was followed up annually for 16 years by teachers and qualified home visitors. The longitudinal study collected a wealth of data in childhood, adolescence, and early adulthood that is now being used to study life course predictors of health and longevity in older adulthood, albeit without data from middle adulthood.

Of the original 6-Day Sample members successfully traced in April 2013 by the NHS Central Register Scotland (88%), 60% were known to be living and 40% were recorded as deceased. This presentation will report on a statistical analysis of data from these 1060 individuals, specifically on the significant factors in early life that predict risk of mortality by 76 years . close to the average life expectancy for the sample's background population. Longitudinal cohort studies have reported significant associations between higher cognitive ability in youth, as well as greater educational and socioeconomic status attainment, and the reduced risk of all-cause and cause-specific mortality. The 6-Day Sample study contributes to this literature. However, while many such studies either began in early adulthood (i.e. at the time of military conscription), or, did not have the initiative and/or resources to conduct intensive data collections in the childhood years, the present study is rare in its ability to explore a wider range of childhood and early adulthood factors, in relation to longevity, at the population level. As well as cognitive ability and socioeconomic status variables, the data include: personality characteristics, home conditions, school attendance and performance, parental health, social and physical activities, childhood illnesses, and nutritional status.

The findings from this life course study may contribute towards shaping new public health interventions, targeted in younger age groups, with the ultimate aim of reducing disease prevalence and extending lifelong health and wellbeing, in our ageing populations.

Corresponding author email: I.Deary@ed.ac.uk

SUDEP and all cause mortality in childhood onset epilepsy

Thomas, R, Swansea University
Lacey, A, Swansea University
White, C, Welsh Epilepsy Research Network
Rees, M, Swansea University

The secure anonymised information linkage (SAIL) project combines a number of health and social health datasets from Wales (population ~3 million). We interrogated primary care datasets to identify people with epilepsy who were prescribed at least one AED (anti epileptic drug) at some point in their life who were diagnosed before the age of 19. We were also able to identify if they had died, and if the death occurred after 2003 we could link to ONS death certificate information.

We scrutinised the primary care databases to look at recorded medical conditions and devised an algorithm that utilised diagnoses that are associated with a symptomatic epilepsy, as opposed to a generalised epilepsy. These features included intellectual disability, traumatic brain injury, hypoxic ischaemic encephalopathy and intercerebral haemorrhage. We used a modified version of the Langan et al [1] definition of a probable SUDEP death from interrogating death records. The adult definition included a cardiac death with epilepsy listed as a comorbidity; but we thought that this was not appropriate in a paediatric population.

14,774 people with epilepsy were identified in the SAIL dataset; which accounts for a total of 352,207 epilepsy years. 930 deaths were identified; 239 of these (26%) were since 2003 and so had linked mortality records. Of these linked deaths, 26 were identified as possible SUDEPs (10%) using the narrow definition. The lowest incident mortality rate was 1.2/1000 epilepsy years for those whose epilepsy started between 6 and 11 and 12 to 18 while the highest was those diagnosed between 0 and 2, 1.6/1000 years. The group (mostly with a probable symptomatic cause of their epilepsy) had 100 total deaths, 37.4/1000 years - which is 16 times greater than the total rate for the cohort.

We conclude that survival is reduced for infants diagnosed with epilepsy who predominantly have a symptomatic epilepsy and school aged epilepsy onset has the best prognosis with SUDEP being rare in school age children. The data presented supports targeting discussions about mortality (expected and unexpected deaths) to children whose epilepsy began before the age of two.

[1] Langan Y, Nashef L, Sander JW. Certification of deaths attributable to epilepsy. J Neurol Neurosurg Psychiatry. 2002;73(6):751-2.

Corresponding author email: a.s.lacey@swansea.ac.uk

Child Medical Records for Safer Medicines (CHIMES).

Helms, PJ, Child Health, University of Aberdeen
Gordon, S, Child Health, University of Aberdeen. This abstract is submitted on behalf of the CHIMES research group.

Background
Record Linkage of Scottish NHS data using the individual Community Health Index (CHI) identifier offers opportunities to address key population health questions. The Child Medical Records for Safer Medicines (CHIMES) programme seeks to determine:

- the acceptability of data linkage to children, parents/guardians and health care professionals.
- the completeness of linked national datasets and their ability to provide comprehensive, accurate, timely and relevant health information
- methods for signal generation in the early detection of potential adverse drug reactions (ADRs) in children.

Methods
Acceptability was assessed with qualitative/quantitative techniques including interviews, content setting focus groups, online questionnaire, and Delphi Survey.
Completeness and utility of linked Scottish NHS data including the Scottish, national Prescribing Information System (PIS) using CHI was assessed by comparing retrospectively assembled virtual exposure cohorts with well established ADR drug profiles and disease patterns. The potential of early medication discontinuation, switching or dose reduction for signal generation was also assessed.

Results
Although young people and parents/guardians had a limited understanding as to how routinely collected NHS data were used they had an expectation that pharmacovigilance would be among these. Opportunities to participate were considered important with the majority of parent/guardians and young people disagreeing with "opt out" as the default position.
HCPs identified a range of problems and concerns including adherence to relevant legal and ethical standards and appropriate governance to ensure confidentiality for both patients and prescribers.
CHI completeness was >90% on dispensed items from January 2010 with current levels reaching 95%. Routinely collected data for insulin prescribing was comparable with published Type 1 diabetes diagnoses in Scottish, UK and Swedish populations. Asthma prevalence rates and prescribing patterns in a virtual cohort were comparable with a concurrent prospective birth cohort and PIS linked data reproduced published findings on the likelihood of hospitalisation for gastrointestinal complications in patients newly prescribed NSAIDs.
Potential ADR signals based on early discontinuations, switching and hospital admissions were found for the obesity drug Orlistat and for antidepressants.

Conclusions
Routinely acquired prescribing data in Scotland can be linked to other NHS data sets with CHI and used to assess medication prescribing profiles and identify potential ADRs.
More work needs to be done to address the issues of consent and assent relevant to children and young people to build public confidence and to demonstrate the benefits of using linked data for improved medicine safety and health gain.

Corresponding author email: p.j.helms@abdn.ac.uk

Important issues with data linkage: A consensus seeking exercise

Hopf, YM, Academic Primary Care, University of Aberdeen
Bond, CM, Academic Primary Care, University of Aberdeen
Francis, JJ, School of Health Sciences, City University London
Haughney, J, Academic Primary Care, University of Aberdeen
Helms, PJ, Child Health, University of Aberdeen

Introduction

Adverse drug events are a major cause of patient safety incidents. Current systems of pharmacovigilance under-report adverse drug reactions (ADRs), especially for paediatric medicines. This is of particular concern due to the widespread use of off-label medicines in children and their resultant increased vulnerability to ADRs. The inclusion of the community health index in the recording of all NHS contacts in Scotland provides the opportunity to link data and thereby identify ADRs by linking prescribing and health utilization data.

Aim

The aim of this study was to seek consensus among Health Care Professionals (HCPs) on potential issues with a planned data linkage and to inform the optimal system design.

Methods

As part of the CHIMES programme (Children's health records for safer medicines) a three round Delphi survey was sent out to a random sample of HCPs (nurses, pharmacists and doctors with an interest in paediatric medicine) in Scotland. The survey was constructed using the Theoretical Domains Framework of Behaviour Change and was informed by the findings of earlier qualitative work. Ethical approval was granted by the North of Scotland Research Ethics Service.

Results

The first round of the Delphi study included 21 questions inviting comments and generated over a 1000 individual statements from 61 participants. These were reduced to 149 items for the second round in whom participants were asked to rate their agreement. After the third round, the retained consensus items focused on professional standards, requirements for linkage and the use and form of potential feedback. Three key requirements of system design were identified namely adherence to current legal and ethical standards, support for HCP time from higher level stakeholders in the NHS and associated employers, and central support for for maintaining accurate and timely data linkage.

Discussion

HCPs indicated their confidence in and intention to facilitate the proposed data linkage as long as clear research governance conditions were adhered to.

Corresponding author email: y.hopf@abdn.ac.uk

Development of an ethics and data linkage training workshop

Flack, F, PHRN, Telethon Institute for Child Health Research
Tan, K, PHRN, Telethon Institute for Child Health Research
Allen, J, University of Western Australia

Introduction

The Australian Government has invested heavily in data linkage, recognising the central role it will play in Australia"s research environment in the future. This expansion of data linkage infrastructure will increase the attractiveness of linked data to many researchers as the availability and accessibility of data collections improves. Data linkage research can provide enormous benefit to our community but it usually requires the use of personal information without consent. Therefore research projects using linked data will generally require review by a HREC. A survey of Australian data linkage units and human research ethics committees (HREC) by the Population Health Research Network (PHRN) identified a need for specialised training in data linkage for members of HRECs.

In response to the combined feedback from data linkage units and HRECs the PHRN developed a training workshop for HREC members. The objectives of the training are to:

- describe the process of data linkage;
- identify the ethical issues specific to research using linked data;
- identify the "best practice protocol" related to the use of linked data and its governance; and
- improve decision making in reviewing research applications proposing the use of linkable data.

The workshop process is designed to support effective adult learning with an emphasis on interactive learning and skills practice.

Results

Workshops have been held across Australia. A total of 114 people representing 37 different HRECs have attended the training. Only 8% of questionnaire respondents had attended any training on data linkage previously and only 29% reported feeling confident to review data linkage applications. More than 90% of participants were satisfied with the workshop structure and organisation and facilitator characteristics and styles. Overall there were low levels of understanding and knowledge of data linkage and the associated legal and ethical issues prior to workshop participation and a significant increase after workshop participation.

Conclusion

The results of the survey of HRECs and workshop evaluation identified that HRECs have a need for training opportunities and an enthusiasm to attend training when it is available. The workshop evaluation demonstrated that the training was able to increase levels of understanding and knowledge about data linkage and the associated legal and ethical issues. We believe this is a good model for the development and delivery of high quality training for HRECs that could be adopted by other groups with expertise in relevant areas.

Corresponding author email: felicity@ichr.uwa.edu.au

Data linkage in a federated system - opportunities and challenges

Dickinson, T, Australian Institute of Health and Welfare

Health care in Australia is provided under a federated model, with the Australian government responsible for funding primary care, and the eight states and territories funding the majority of acute and sub-acute care. This means that data about patient care is under the custodianship of a number of different government entities - each having different provisions in place to guide the release of data for research. For researchers there is often considerable interest in understanding patient experience across the entire health system over time - achieving this therefore requires linking data from Commonwealth and state and territory sources.

Recently the Australian government introduced new arrangements under which it will make its data available. The Commonwealth framework:

- Is based on a number of underlying principles
- Stipulates the legislative, technical, organisational and governance requirements for undertaking data linkage work, including the requirement that work deemed high risk be undertaken only by an accredited integrating authorities
- Specifies roles of data custodians, users and integrating authorities
- Describes arrangements for obtaining accreditation to function as an integrating authority

States and territories also have arrangements in place for releasing their data for research projects involving linkage. While these are largely complementary with the Commonwealth approach, there are areas where challenges arise in meeting the requirements of all parties.

The Australian Institute of Health and Welfare is currently one of only two accredited integrating authorities in Australia. This paper will describe the process of obtaining accreditation as well as the challenges of adopting the new arrangements in a federated environment.

Corresponding author email: teresa.dickinson@aihw.gov.au

Using a data system used to store blood results Scottish Care Information Gateway (SCI) Store in healthcare research

McAllister, DA, University of Edinburgh
Hughes, KA, University of Edinburgh
Lone, N, University of Edinburgh
Mcknight, J, NHS Lothian
WIld, SH, University of Edinburgh

Introduction

Hyperglycaemia in people without diabetes is common following admission to hospital with acute illness. Estimates of the risk of developing diabetes among this group are based on small studies (maximum of 630 patients with most having less than 200 patients) and of limited duration of follow-up (up to 3 months). As such, the longer-term risk of diabetes by admission glucose level, which would inform decisions regarding follow-up testing, is unknown.

We aimed to link large routine datasets recording hospital admissions, glucose blood results and diagnosis with diabetes in order to estimate the 5-year risk of diabetes for all emergency hospital admissions according to blood glucose level. We further aimed to estimate this risk according to age, sex, deprivation and healthcare specialty (ie medical and surgical).

Methods

We used an existing linkage between the Scottish diabetes register (Scottish Care Information (SCI)-Diabetes) and Scottish hospitalisation data to identify the 5-year risk of developing diabetes among patients admitted to hospital for medical and surgical emergencies.

We performed a novel linkage of the investigation results database (SCI-Store) held within NHS Lothian to this existing linked dataset to obtain glucose results. NHS Information Services Division (ISD) provided CHI-numbers and pseudonymised codes for patients with emergency admissions to NHS Lothian. The NHS Lothian SCI-store manager extracted glucose results and provided these which were linked to the original data using pseudonymised codes.

Results

In the initial linkage 22,403 patients without diabetes were identified, of whom 745 were subsequently diagnosed with diabetes within 5 years.

We used a logistic regression model to obtain predictions of the 5-year risk of diabetes according to glucose level, using squared and cubic terms to allow for non-linearity. The risk was ⩽1% for glucose levels from 2 to 4 mmoll-1. The risk increased broadly linearly to 9.6% (95%CI 8.5-10.8) for a glucose level of 10 mmoll-1.

Discussion

Linkage to SCI-store presents challenges but, even on a small subset of our final dataset, enabled us to improve on current estimates of the 5-year risk of diabetes among patients admitted to hospital for medical and surgical emergencies according to glucose level.

We intend to obtain data for a larger sample of patients in order to allow us to estimate risk stratified by age, sex, deprivation, medical/surgical specialty and in sub-groups with specific diagnoses.

Corresponding author email: david.mcallister@ed.ac.uk

Combining health, physical function and social care data - a multidatabase linkage project

McGilchrist, M M, University of Dundee
Donnan, P T, University of Dundee
McMurdo, M E, University of Dundee
Frost, H, Scottish Collaboration for Public Health Policy and Research
Goodbrand, J, University of Dundee
Cochrane, L, University of Dundee
Witham, M D, University of Dundee

Background:
Effective care of older people requires attention to multiple dimensions of care, including health problems, physical and psychosocial function, and social support. The current drive towards closer integration of health and social care services reflects this. Despite this, there is a lack of integrated data that seeks to combine healthcare data with social care data. Healthcare outcomes can be confounded by functional status in older people, and many health analyses are weakened by a lack of such functional data. Similarly, social care data analyses need to incorporate detailed healthcare data if meaningful results are to be derived.

Process description:
Combining health and social care data requires working across different organisational cultures, negotiating different bureaucracies, and understanding data sources collected in distinct ways for distinct purposes. We describe the results of a project that has successfully linked data from three sources: 1) Healthcare data held by the Health Informatics Centre, 2) activities of daily living data on admission and discharge from the Dundee Medicine for the Elderly rehabilitation service, and 3) social care data from Dundee Social Services.

Health and social care data on approximately 30,000 individuals aged 65 and over seen by Dundee Social services over the past 20 years were linked using seeded pseudo-CHI numbers, and combined with data from 5500 similarly linked rehabilitation admissions that provided functional data. The combined, anonymised datasets are held within the Tayside Academic Health Sciences Consortium safe haven.

Outputs:
Barriers and facilitators to the successful retrieval and combining of these datasets will be discussed, along with examples of how the data is facilitating novel analyses: 1) effects of bisphosphonate prescribing on rehabilitation outcomes, 2) effect of allopurinol prescribing on hip fracture, 3) the impact of functional status on outcome in chronic kidney disease, 4) prediction of future care home admission using both health and social service data predictors. We will also describe dissemination plans, plans to promote of multiagency involvement and plans for future use of this novel dataset.

Corresponding author email: m.m.mcgilchrist@dundee.ac.uk

Linkage of UK National Liver Transplant Registry and Hospital Episode Statistics: Methods and Initial Validation

Tovikkai, C, Department of Surgery, University of Cambridge; Clinical Effectiveness Unit, Royal College of Surgeons of England

Charman, SC, Clinical Effectiveness Unit, London School of Hygiene and Tropical Medicine

Praseedom, RK, Department of Surgery, University of Cambridge

Gimson, AE, Liver Transplant Unit, Addenbrooke's Hospital

Watson, CJE, Department of Surgery, University of Cambridge

van der Meulen, JHP, Clinical Effectiveness Unit, London School of Hygiene and Tropical Medicine

Introduction

The UK national liver transplant registry (UKT) has been established since 1994 and contains clinical information about recipient, donor and operative characteristics related to liver transplantation in eight transplant centres in the UK and Ireland. In order to expand the research potential of the UKT database, we undertook the linkage of UKT with the Hospital Episode Statistics (HES) and an initial validation exercise.

Methods

Linkable records of first liver transplantation between April 1997 and March 2010 from each database were identified. The linkage was carried out based on deterministic linkage criteria, using NHS number, gender, date of birth and postcode. Matched records were considered successfully linked if the date of transplant in the two sources differed by no more than one day. The final linked dataset was validated in terms of the agreement of indication for transplantation in the two databases. Diagnosis codes in HES, which were coded in the tenth revision of the International Classification of Diseases (ICD-10), were translated into a previously validated disease classification for use in liver transplantation by a consensus group of independent experts, two transplant surgeons and a transplant hepatologist. The agreement of the indication as recorded in UKT with the indication as translated from HES was assessed by the proportion of agreement and the kappa coefficient. Survival rates by each indication group as determined by UKT and HES were also compared.

Results

There were 5,815 linkable records in the UKT database, of which 4,959 records were successfully linked with HES (85.3%). Amongst these successfully linked records, 4,922 records (99.3%) had at least one diagnosis coded in HES that was relevant to an indication for liver transplantation. The overall proportion of agreement for indications coded in UKT and those translated from HES was 77.8% (95% CI: 76.6-79.0). The kappa coefficient was 0.75 (95% CI: 0.74-0.76). Kaplan-Meier survival curves of each indication group that were coded in UKT and translated from HES showed no difference.

Discussion

As far as we are aware, this is the first time the UKT database has been linked to HES. This linked database will be a useful resource for research in liver transplantation. Having demonstrated good agreement of the primary diagnosis, we suggest that, in countries where a transplant registry is not available, administrative data records using the ICD-10 classification system may be useful for transplant audit and research.

Corresponding author email: ct422@cam.ac.uk

Area and individual socioeconomic factors and cancer risk: a population cohort study in Scotland

Sharpe, KH, Information Services Division, NHS National Services Scotland
McMahon, AD, University of Glasgow, College of Medical, Veterinary and Life Sciences: Dental School
Raab, G, University of St Andrews
Brewster, DH, Public Health Sciences, Edinburgh University Medical School
Conway, DI, University of Glasgow, College of Medical, Veterinary and Life Sciences: Dental School

Background: Socioeconomic inequalities in cancer risk differ by tumour site, age and sex. Lung and upper aero-digestive tract (UADT) cancer risk contribute 90% (males) and 81% (females) to cancer risk social inequalities in Scotland and are associated with low socioeconomic circumstances. We investigate the relative association with cancer risk of country of birth, marital status, area deprivation and individual socioeconomic variables (economic activity, education level, occupational social class, car ownership, housing tenure) for lung, UADT and all cancer risk (excluding non melanoma skin cancer).

Methods: We linked Scottish Longitudinal Study, vital event registries and Scottish Cancer Registry data and followed 203 658 cohort members aged 15+ years from 1991- 2006. We calculated relative risks and 95% confidence intervals using fully adjusted Poisson regression models for each sex offset by person-years of follow-up.

Results: 21 832 first primary tumours, including 3 505 lung and 1 206 UADT tumours were diagnosed corresponding to 3.05 million person-years of follow-up. For females, car ownership and housing tenure were more strongly associated with increased risk. For males unemployment was consistently associated with increased cancer risk (except lung), while education was not associated with increased risk. For lung cancer, area deprivation remained significant even after adjustment for individual variables in both sexes, suggesting the area affect can not be fully explained by individual socioeconomic circumstances. Finally, being born in Scotland, divorced or widowed was associated with increased risk regardless of sex.

Conclusion: Different and independent socioeconomic variables are associated with different cancer risks in different sexes.

Corresponding author email: katharine.sharpe@nhs.net

Socio-economic patterning in early mortality of patients aged 0-49 years diagnosed with primary bone cancer in Great Britain, 1985-2009

Blakey, K, Institute of Health and Society, Newcastle University, Newcastle-upon-Tyne, England, United Kingdom
Feltbower, RG, Paediatric Epidemiology Group, University of Leeds, Leeds, England, United Kingdom
Parslow, RC, Paediatric Epidemiology Group, University of Leeds, Leeds, England, United Kingdom
James, PW, Institute of Health and Society, Newcastle University, Newcastle-upon-Tyne, England, United Kingdom
Stiller, C, Childhood Cancer Research Group, Department of Paediatrics, University of Oxford, England, United Kingdom
Norman, P, School of Geography, University of Leeds, Leeds, England, United Kingdom
Gerrand, C, Northern Institute for Cancer Research (NICR), Newcastle University and North of England Bone and Soft Tissue Tumour Service, Newcastle-Upon-Tyne, England, United Kingdom
McNally, RJQ, Institute of Health and Society, Newcastle University, Newcastle-upon-Tyne, England, United Kingdom

Purpose: The study aim was to explore socio-economic patterning in early mortality rates for osteosarcoma and Ewing sarcoma. Early mortality was considered as a proxy for delayed diagnosis.

Methods: All cases of osteosarcoma and Ewing sarcoma aged 0-49 years, diagnosed during 1985-2009 were included. Data were provided through regional cancer registries. Logistic regression models analyzed the odds of early mortality at fixed periods of 3, 6 and 12 months. Associations with Townsend deprivation score and its components (percentage of households over-crowded, non-car ownership, non-home ownership, or residents unemployed) were examined at small-area level (census ward for England and Wales, postcode sector for Scotland). A time series of Townsend deprivation scores was constructed by apportioning the four constituent measures from the 1991 and 2001 censuses (applied to 1985-1995 and 1996-2009 data, respectively) to the 2001 census geography. Odds ratios (ORs) and 95% confidence intervals (CIs) are presented. Statistical significance was taken to be P < 0.05.

Results: The study analyzed 2562 osteosarcoma cases aged 0-49 years (820, 1262 and 480 aged 0-14, 15-29 and 30-49 years respectively) and 1711 Ewing sarcoma cases (670, 822 and 219 aged 0-14, 15-29 and 30-49 years respectively).

For osteosarcoma, after adjustment for age, mortality at 3, 6 and 12 months was not significantly linked with Townsend deprivation (P=0.253, 0.252 & 0.148 respectively). However it was significantly greater in areas of higher unemployment (OR= 1.06, (CI 1.02, 1.11; P=0.018); OR = 1.04 (CI 1.01, 1.08; P=0.020), OR = 1.04 (CI 1.02, 1.06; P=0.001) per 1% increase in unemployment respectively). For Ewing sarcoma there were no significant associations between mortality and Townsend deprivation score at 3, 6 and 12 months after diagnosis (P=0.996, 0.134 & 0.200 respectively), nor with any of its four components.

Conclusions: This study has revealed an important finding. For osteosarcoma, early mortality was associated with residence in areas of higher unemployment. This finding suggests that delays in diagnosis of osteosarcoma may be socio-economically determined.

Corresponding author email: karen.blakey@newcastle.ac.uk

Can surveys provide a means of providing representative information on cancer survivors - a data linkage study

Atherton, I, University of Stirling (Highland Campus)
Dibben, C, University of St Andrews
Evans, J, University of Stirling (Stirling Campus)
Hubbard, G, University of Stirling (Highland Campus)

Background - Prevalence of cancer survivors in Scotland, like many countries, is increasing. People are living longer lives and consequently are at higher risk of developing cancer. Once diagnosed, medical care has progressed such that prognosis has improved considerably over recent years. The importance of well-being for this rising proportion of the population is a notable public health concern. Gaining insight into cancer survivors' wellbeing requires representative data. In this study we investigate the implications of using self-reported survey data to identify cancer survivors in comparison to cancer registry data.

Methods - Data from the 1995, 1998 and 2003 rounds Scottish Health Survey, a nationally representative cross-sectional survey, were pooled into a single cross-sectional dataset. Respondents were asked in the survey if they had a long-term illness. Those answering in the positive were then able to identify what the condition was, cancer being one of the possible responses recorded. The survey data was then linked to cancer registry data for respondents who agreed to linkage (over 90%) to identify those who were recorded as being cancer survivors. Three groups were thus created from the linked self-reported survey information and cancer registry data linked data, specifically categorising respondents as: (1) never diagnosed with cancer on cancer registry, (2) diagnosed and self-reporting to have cancer, and (3) diagnosed but not self-reporting cancer. We compared these three groups with regards to self-assessed health, self-reported reductions in recent activity, and psychological morbidity (GHQ-12) using logistic regression models.

Findings - 486 cancer survivors were identified from the linked cancer registry data. Of these, only 299 (38.5%) indicated themselves as having had cancer. Logistic regression models demonstrated this group, in comparison to those never diagnosed with cancer, more likely to have psychiatric morbidity (adjusted OR 2.42 CI 1.75-3.35) to have had reduced activity in the two weeks prior to survey (adjusted OR 2.74 CI 2.02-3.72) and to have worse self-assessed health (adjusted OR 4.27 CI 3.08-5.92). Conversely, cancer survivors who did not indicate themselves to have cancer were no more likely than those never diagnosed to have psychiatric morbidity or reduced activity. However, they also were more likely to indicate poor self-assessed health (adjusted OR 1.34 CI 1.05-1.72).

Conclusions - The study demonstrates the value of data linkage to understanding the wellbeing of people who have experienced major health events. Reliance on surveys alone risks missing a notable proportion of cancer survivors.

Corresponding author email: iain.atherton@stir.ac.uk

Linkage of primary care, pharmacy invoice and hospital admissions records to a national joint registry: the PRESS-UP Cohort.

Allepuz, A, Catalan Arthroplasty Registry (RACat), AIAQS, Government of Catalonia (Generalitat de Catalunya)
Martinez-Cruz, O, Catalan Arthroplasty Registry (RACat), AIAQS, Government of Catalonia (Generalitat de Catalunya)
Alves, L, USR Girona, Idiap Jordi Gol
Garcia-Gil, M, USR Girona, Idiap Jordi Gol
Espallargues, M, Catalan Arthroplasty Registry (RACat), AIAQS, Government of Catalonia (Generalitat de Catalunya)
Ramos, R, USR Girona, Idiap Jordi Gol
Prieto-Alhambra, D, NDORMS, University of Oxford, and USR Barcelona, Idiap Jordi Gol

PURPOSE: Use of non-steroidal anti-inflammatory drugs (NSAIDs) in the first year following total knee (TKA) and hip (THA) arthroplasty is a surrogate for pain and therefore a potential predictor of implant survival. However, national joint registries hold no reliable data on drug utilization.

We studied the association between NSAID utilization and implant survival in the PRESS-UP cohort: Catalan Arthroplasty Registry (RACat) data linked to primary care records and pharmacy invoice gathered in the SIDIAP Database (www.sidiap.org).

METHODS

Participants: patients aged >=40 years undergoing TKA/THA for knee/hip osteoarthritis in RACat (2005-July/2012), who could be identified in SIDIAP using trusted third party linkage. Patients receiving revision surgery in the first year post-surgery were excluded.

Exposure: NSAID utilisation was quantified in number of Daily Defined Doses (DDDs) according to the WHO ATC/DDD index, and categorized into quintiles.

Outcome and analysis: Implant survival (in years) was the main outcome. Fine and Gray regression was used to estimate sub-hazard ratios (SHR) according to NSAID utilisation, adjusted for: age, gender, socio-economic status, Charlson Comorbidity Index, alcohol drinking, smoking status, and body mass index.

RESULTS: 23,197/36,897 (62.9%) TKA and 16,703/29,665 (56.3%) THA patients registered in RACat were linked to SIDIAP. 22,221/23,197 (95.8%) and 10,173/16,703 (60.9%) participants underwent surgery for osteoarthritis. TKA and THA participants were followed up for a median (inter-quartile range) of 3.20 (2.08-4.71) and 2.22 (1.17-3.72) years respectively. In this time, 724 (3.3%) TKA and 428 (2.6%) THA patients were revised, 634 (2.9%) and 2,105 (12.6%) died, and 83 (0.4%) and 108 (0.7%) were lost to follow-up. Rates of revision after the first year were positively associated with utilisation of NSAIDs during the first year: adjusted sub-hazard ratio (SHR) 1.22 [95% Confidence Interval 1.14-1.31; p<0.001 for trend] per quintile for TKA, 1.29 [1.08-1.54; p=0.005 for trend] for THA.

CONCLUSIONS: NSAID utilisation in the first year following elective TKA/THA for knee/hip osteoarthritis is directly related to revision risk in subsequent years. This is an interesting early surrogate, which could be used both in clinical practice and in research studies. Linkage of primary care records, pharmacy invoice and registry data can improve the identification of predictors of different outcomes in specific relevant subpopulations.

Corresponding author email: lalves@idiapjgol.info

Incidence of and risk factors for Motor Neurone Disease in UK women: a prospective study

Doyle, P, London School of Hygiene & Tropical Medicine
Brown, A, University of Oxford

Background: Motor neuron disease (MND) is a severe neurodegenerative disease with largely unknown aetiology. Most epidemiological studies are hampered by small sample sizes and/or the retrospective collection of information on behavioural and lifestyle factors.

Methods: The Million Women Study (MWS) has been linked to Hospital Episode Statistics for England (HES) and to Scottish morbidity records (SMR) for Scotland. 1.3 million participating women in middle age were followed up for incident and/or fatal MND using NHS hospital admission and UK mortality data. Women were classified as having MND if they had a hospital admission record with an ICD10 code of G12.2 and/or a death registration with any mention of G12.2. Completeness of case-finding was assessed by comparison with general practitioner (GP) records for a subset of women. Lifestyle, behavioural and other factors were obtained prospectively from a self-administered questionnaire at study entry. Adjusted relative risks were calculated using Cox regression models.

Findings: After an average follow-up of 9.2 years, 752 women had a new diagnosis of MND. In random samples of (a) cases identified via hospital records, 91% had their diagnosis confirmed by their GP, and (b) women with no hospital record of MND, 100% of GPs confirmed they had no record of a diagnosis of MND. Age-specific rates of MND increased with age, from 1.9 (95% CI 1.3 - 2.7) to 12.5 (95% CI 10.2 - 15.3) per 100,000 women aged 50-54 to 70-74, respectively. There was no significant variation in risk of MND with region of residence, socio-economic status, education, height, alcohol use, parity, use of oral contraceptives or hormone replacement therapy. Ever-smokers had about a 20% greater risk than never smokers (RR 1.19 95% CI 1.02 to 1.38, p = 0.03). There was a statistically significant reduction in risk of MND with increasing body mass index (p for trend=0.009): obese women (body mass index, 30 kg/m2 or more) had a 20% lower risk than women of normal body mass index (20 to <25 Kg/ m2)(RR 0.78 95% CI 0.65-0.94; p = 0.03). This effect persisted after exclusion of the first three years of follow-up.

Interpretation: MND incidence in UK women rises rapidly with age, and an estimated 1 in 575 women are likely to be affected between the ages of 50 and 75 years. Smoking slightly increases the risk of MND, and adiposity in middle age is associated with a lower risk of the disease.

This abstract extends information published in BMC Neurology 2012, 12:25 doi:10.1186/1471-2377-12-25

Corresponding author email: anna.brown@ceu.ox.ac.uk

NOAH'S ARK: A GLOBAL DATABASE SUCCESS STORY OR WHAT MIGHT ONE LEARN AT A ZOO?

Harvey, P,

This presentation compares and contrasts the information management strategy of the global captive animal management community with that of the National Cancer Intelligence Network (NCIN) in England.

It will:

- highlight the potential risks to quality in terms of both process and content resulting from the current NCIN information management strategy and
- clarify the lessons to be learned from the International Species Information Systems (ISIS) approach - with a particular focus on improved efficiency and data quality

The presentation considers the data reporting requirements for a hospital in England caring for a child diagnosed with a brain tumour. Data pertaining to that child are sent by that hospital to the National Cancer Registry, to the Children's Cancer Registry, the National Brain Tumour Registry, as well as for Cancer Waiting Times management. Data will also be submitted for Health Episodes Statistics, National Clinical Audit, and various cancer treatment datasets. In many cases, data might also be submitted to a Clinical Trial Unit and/or Biobank.

The modern Zoo and Aquarium network provides a safety net for the survival of endangered species. Successful delivery of cancer care is as dependent on effective inter-institutional access to accurate data as successful captive animal management.

These two industries share a convergent purpose to provide "sophisticated knowledge management tools and connection to a global professional scientific network" but employ divergent strategies to achieve this aim.

The presentation looks at three questions:

1. What, if anything, is wrong with the way cancer data are handled?
2. Which strategy more efficiently and effectively supports the shared goal?
3. What, if anything, might cancer information, service and research networks learn from the experiences of the zoo world?

Corresponding author email: pamela.b.harvey@gmail.com

Big data exploration: Development of a novel metadata format to record longitudinal study consent models for record linkage
Abstract for a poster presentation

McMahon, C, MRC Centre of Epidemiology for Child Health, UCL
Dezateux, C, MRC Centre of Epidemiology for Child Health, UCL
Kehoe, D, AIMES Grid Services Ltd.
Castillo, T, MRC Centre of Epidemiology for Child Health, UCL

Background

Longitudinal studies can produce big data defined as data of a volume and/or complexity requiring infrastructure and querying techniques beyond that provided by standard tools. The longevity of this data is such that it must be matched with consent models equally as enduring. However, developing these models is a challenging process hindered by a lack in standardised methods of recording them.

Longitudinal studies, in addition to others, can be structured into sets of processes each connecting to stages within the research data lifecycle. One of the early processes in the lifecycle is the requesting of consent to participate in the study and to share personal records for analysis and record linkage endeavours.

The aim is to develop and demonstrate a method of recording longitudinal study consent models in a novel metadata (data about data and associated processes) format. This method will reflect the need to record vital information about the use and linkage of personal records.

Methods

The consent forms studied belonged to the following studies: ALSPAC, BHPS, Health Survey for England, Millennium Cohort Study, Scottish Health Survey, UK Biobank, and Understanding Society. The identified concepts were combined and became the basis of a class diagram.

Development of the metadata format involved an exploration into the use of object oriented formalisms as supported by object oriented analysis and design. Object orientation is a paradigm for organising information and associated processes into manageable subunits. These techniques facilitated the simplified expression of identified concepts and assisted model harmonisation efforts.

The diagram was then mapped to elements within the metadata standard, Data Documentation Initiative v3.1 (DDI-L). DDI-L is an internationally recognised metadata standard that encodes metadata using XML; the cyclical nature of DDI-L complements the research data lifecycle. This mapping exercise proved an opportunity to test whether the initial harmonised consent model could be mapped to a pre-existing metadata standard and to help identify areas of improvement.

Interpretation

To date, development of the initial model has been successful as have efforts to map the model to DDI-L. Initial results demonstrate a metadata format enabling the recording of consent model metadata which will help researchers establish the extent to which consent has been given and thus the degree to which record linkage may occur. Future work involves further developing the model and possibly extending DDI-L to enhance mappings between the model and the standard.

Acknowledgements

MRC CASE studentship

Corresponding author email: christiana.mcmahon.11@ucl.ac.uk

Real World Oncology Evidence - use of a national systemic anti-cancer therapy dataset to improve outcomes, uptake and value for Scotland

Moore, L, Roche Products Limited
Mcnamara, L, Roche Products Limited
Stevenson, G, Roche Products Limited

Background: Cancer represents a significant management challenge for the health system. More than 1 in 3 people in Scotland will develop some form of cancer during their lifetime and will rely on a range of interventions, including pharmacological treatments, in order to improve their outcomes and their quality of life. A database that collects granular information on the utilisation of cancer medicines presents an opportunity to evaluate the introduction of new treatments into the service. Developments in electronic prescribing and data collection in cancer offers a unique opportunity to understand, develop and sustain access to the right treatments, at the right time based upon real patient outcomes.

Methods: This research aims to clarify both the opportunities presented by the potential availability of an anti-cancer therapy database in Scotland, as well as the gaps to be filled in order to achieve this. Scotland has always been at the forefront of epidemiology and data collection with the Information Service Division of Scottish Government producing key statistics for the population at large.However, accurate information on usage of anti-cancer therapies is not yet available. Scotland is uniquely placed in having a national strategy on electronic prescribing and record keeping (CEPAS) and opportunities in data collection in cancer therefore exist. The potential role of complete national dataset to improve outcomes, uptake and value for Scotland have been assessed through the use of an analogue of a similar dataset and electronic prescribing system within a Provider Trust in England.

Results: Recent joint working between Roche Products Ltd. and a Provider Trust in England has shown that data being collected on chemotherapy treatment has significant value for NHS, Academic stakeholders and industry when used to its full potential. Better utilisation of data would allow NHS Scotland to evaluate the introduction of new treatments, benchmark quality of service, support the rapid identification of those suitable for clinical trials and help manage complex patient access schemes required to ensure best value to NHS Scotland. Realising this value requires a coordinated approach across the health system as well as supporting infrastructure.

Conclusion: Improving access, improving equity, improving efficiency and improving outcomes all fit with current NHS Scotland policies on Quality, Efficiency and Productivity, and the Cancer Action Plan. However, currently the NHS does not know how many cancer patients actual access medicines that are aimed at improving disease outcomes. Equally, the NHS does know the outcomes from the medicines it routinely funds. Utilising existing real world data, systems and processes to realise this clarity and drive improvements will benefit patients and help the NHS to achieve its priorities. To enable the use of data in such a way NHS Scotland should look to move forward with improved utilisation of the existing data collection mechanisms to create a single data set on the use of cancer medicines.

Produced by Roche Products Limited, April 2013
RXUKECON00010b

Corresponding author email: gregor.stevenson@roche.com

An Evaluation of NPIRS - Ireland's National Psychiatric Inpatient Reporting System

Moran, R, Health Research Board1
Donal McAnaney2, Richard Wynne2, Brendan Curran1, Breda Naddy1, Antoinette Daly1, Sarah Craig1, D, Work Research Centre 2

An independent evaluation of NPIRS - Ireland's National Psychiatric Inpatient Reporting System was carried out. An overall accuracy rate of 97% was found which is high by international standards. The primary admission and discharge diagnosis fields had the lowest accuracy scores (94%, 93%). Completeness rates were also very high (99%). Stakeholders rated NPIRS highly on a number of dimensions including relevance. NPIRS data was used for a variety of purposes including policy, planning, monitoring and research. Further improvements in NPIRS are highly dependent on external stakeholders and data providers.

Corresponding author email: rmoran@hrb.ie