

See discussions, stats, and author profiles for this publication at:
<https://www.researchgate.net/publication/236947335>

Developing a survey instrument to assess the readiness of primary care data, genetic and disease registries to conduct...

Article in *The Journal of Innovations in Health Informatics* · November 2013

DOI: 10.14236/jhi.v20i3.26 · Source: PubMed

CITATIONS

3

READS

110

11 authors, including:



Simon de Lusignan

University of Surrey

410 PUBLICATIONS 4,374 CITATIONS

SEE PROFILE



Georgios Michalakidis

University of Surrey

12 PUBLICATIONS 114 CITATIONS

SEE PROFILE



Paul John Krause

University of Surrey

167 PUBLICATIONS 2,701 CITATIONS

SEE PROFILE



Frank michael Sullivan

University of Toronto

323 PUBLICATIONS 5,351 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:



HiCure (<http://sites.birzeit.edu/hicure/>) [View project](#)



PA-SCER is a research study investigating the contribution that Physician Associates make to secondary care health services in England. [View project](#)

Refereed paper

Developing a survey instrument to assess the readiness of primary care data, genetic and disease registries to conduct linked research: TRANSFoRm International Research Readiness (TIRRE) survey instrument

Emily Leppenwell

Hon Research Assistant, Department of Health Care Management and Policy, University of Surrey, Guildford, UK

Simon de Lusignan

Professor of Primary Care and Clinical Informatics, Department of Health Care Management and Policy, University of Surrey, Guildford, UK; Division of Population Health Sciences and Education, St George's – University of London, UK

Marilia Tristan Vicente

Intern/Hon Research Assistant, Division of Population Health Sciences and Education, St George's – University of London, UK

Georgios Michalakidis

Postgraduate Researcher

Paul Krause

Professor of Software Engineering

Department of Computing, University of Surrey, Guildford, UK

Steve Thompson

Hon. Research Assistant, Department of Health Care Management and Policy, University of Surrey, Guildford, UK

Mark McGilchrist

Senior Research Fellow

Frank Sullivan

Professor of Primary Care

Division of Population Health Science, University of Dundee, UK

Terry Desombre

Professor of Health Care Management and Head, Department of Health Care Management and Policy, University of Surrey, Guildford, UK

Adel Taweel

Technical Director and Lecturer in Software Engineering

Brendan Delaney

Professor of Primary Care

Department of Primary Care Research and Public Health Sciences, King's College London, UK

ABSTRACT

Background Clinical data are collected for routine care in family practice; there are also a growing number of genetic and cancer registry data repositories. The Translational Research and Patient Safety in Europe (TRANSFoRm) project seeks to facilitate research using linked data from more than one source. We performed a requirements analysis which identified a wide range of data and business process requirements that need to be met before linking primary care and either genetic or disease registry data.

Objectives To develop a survey to assess the readiness of data repositories to participate in linked research – the Transform International Research Readiness (TIRRE) survey.

Method We develop the questionnaire based on our requirement analysis; with questions at micro-, meso- and macro levels of granularity, study-specific questions about diabetes and gastro-oesophageal reflux disease (GORD), and research track record. The scope of the data required was extensive. We piloted this instrument, conducting ten

preliminary telephone interviews to evaluate the response to the questionnaire.

Results Using feedback gained from these interviews we revised the questionnaire; clarifying questions that were difficult to answer and utilising skip logic to create different series of questions for the various types of data repository. We simplified the questionnaire replacing free-text responses with yes/no or picking list options, wherever possible. We placed the final questionnaire online and encouraged its use (www.clininf.eu/jointirre/info.html).

Conclusion Limited field testing suggests that TIRRE is capable of collecting comprehensive and relevant data about the suitability and readiness of data repositories to participate in linked data research.

Keywords: computerised medical record systems, data collection, diabetes mellitus, general practice, genetic databases, medical informatics, oesophageal diseases, primary health care, registries

Introduction

The European e-Health Action Plan calls for interoperability between different health computer systems so that international data can be collected on the quality of care and for research.¹ Translational Research and Patient Safety in Europe (TRANSFoRm) is a European Union (EU)-funded international research project aimed at delivering part of this agenda.² The aim of TRANSFoRm is to bring together existing primary care databases with genetic and/or cancer registry data repositories. These linked data sets can then be used to conduct more sophisticated translational and biomedical research.

These three data sources, genetic data often in biobanks, cancer registries and routinely collected primary care data, have been widely used for research, but are rarely combined. Genetic data have been increasingly collected into biobanks since completion of the mapping of the human genome between 2001 and 2004;^{3,4} although biological specimens have been used for a longer period in pathology and forensic science.^{5,6} Cancer registries have been in place for some years^{7,8} and more recently have become international with the aim of understanding the epidemiology of conditions,⁹ as well as the effectiveness of

treatments.¹⁰ Primary care clinical recording in the UK has shifted decisively from paper to computer over recent years^{11,12} with routine data widely used for research,¹³ but with relatively little linkage of data beyond often disease-specific programmes in individual localities.

A central component of the TRANSFoRm project is the development of three-dimensional use-cases in two clinical domains: type 2 diabetes and gastro-oesophageal reflux disease (GORD). An epidemiological study of people with type 2 diabetes aspires to link genetic and primary care data sets and set out to explore the risk of complications and responses to oral medications. In the GORD domain there are two use-cases. The first is a case-controlled study of the association of GORD symptoms and the use of acid-blocking drugs (specifically proton pump inhibitors; PPIs) with developing Barrett's disease and cancer of the oesophagus. The second is a randomised controlled trial of on-demand compared with continuous PPI drug use, including monitoring patients' symptoms and outcomes.

We set out to develop a survey instrument that could determine if a database was able to meet the requirements analysis and participate in linked data research projects as set out in these use-cases.

Method

Overview

We report the development of a survey instrument to collect the data identified in our requirements analysis.^{14,15} In brief, this was part of the TRANSFoRm project, whose relevant work tasks are summarised in Table 1. Our requirements analysis identified data requirements at the micro-, meso- and macro levels, study-specific requirements and proposed that the research track record was used as a sensitivity analysis (Figure 1). Functionally, our micro-requirements describe potential study data quality¹⁶ and its readiness to incorporate into linked research; the meso-level items to the technical readiness of the system and data export systems. Key among the macro-sociocultural issues are governance and regulatory readiness, without which it is not possible to proceed. Study-specific information, driven by the nature of the data held and the sample size required to flag readiness for a particular study.

We adopted a two-phase approach to determine the capability of primary care data repositories to link to genotype data repositories and disease registries. Phase one was the development of the survey instru-

ment reported here, and phase two was data collection and analysis.

TIRRE questionnaire content

We created a framework for the data needed to assess the fitness and preparedness of repositories to participate in linked research. The framework drew on principles from interoperability, EPR architecture, modelling and existing research data schema. Using this framework, we developed a questionnaire that could be used to collect data through a telephone survey or be self-administered online. The framework ensured that the survey data were collected at three levels of granularity using standard headings for each section of the questionnaire; additionally collecting the study-specific data. The heterogeneity of the data meant that we needed to extract comprehensive information about the data and how it was structured. The first section, the micro level was concerned with the data source, the data itself, metadata, the potential for linkage or achieving semantic interoperability between data sources,¹⁷ and details of how many studies have been published using the data. The second section, the meso level, explored data extraction,¹⁸ the architecture of the computerised medical record

Table 1 Details of the TRANSFoRm work tasks

Description of work tasks (WT) 6.1 and 6.2 of the TRANSFoRm project:

WT 6.1: Requirements analysis of EHRs

1. Using the results of the EHR capacity study within the EGPRN and ESPCG networks (WT1.1) we will conduct an in-depth study of the most common EHR systems used in Europe to examine the availability of API details. The scope of the study will include patient held records, which may hold substantially less coded and structured data. These will include Microsoft and Google patient record systems – and countries where health cards are used.
2. We will conduct a parallel in-depth study of data repositories that can be used for clinical trials. We will look to identify local, EHR brand-specific and health-system access points to primary care data. The types of data access points we might be able to run queries on include: (1) billing or performance indicator extracts of routine data; (2) sentinel networks or research network database; and (3) national data extract systems with closed API. These may provide pragmatic quick win access to primary care data while a longer term access is being developed.

WT 6.2: Requirements analysis of genotype data repositories

We will conduct a parallel in-depth study of (genotype) clinical data repositories across Europe and their potential use for clinical research. The scope of the study will include structured genotype data and potential integration points with patient health care for biomedical and translational clinical research. Genotype data is normally held by biobanks or other research organisations either as sample identification information or specific codes for single nucleotide polymorphisms (SNPs).

Micro level		Meso	Macro		Study	
Data Source Level		Record System Level	Health System Organisational Level	Socio-cultural Level	Study Use-case specific	
<p>A. Genetic Databases or other (e.g. hospital, social care, cohorts):</p> <ul style="list-style-type: none"> • Focus (general database/specific disease) • Demographics definition (national representative population/inclusion and exclusion criteria) • Only for genetic databases: <ul style="list-style-type: none"> • Links to complete primary care records • Phenotype information • Links to routine health records of patients <p>B. Primary Care Databases or Disease Registries:</p> <ul style="list-style-type: none"> • Central or distributed data store network • Links to genetic databases • Single or multiple EPR vendors (and which) <p>C. For all types of resources:</p> <ul style="list-style-type: none"> • Potential to support studies • Cross-sectional study eligibility • Capability to follow cohorts • Cause of death information • Methods of collection <ul style="list-style-type: none"> • Upload types • Critical trial forms • Aims and Publications (information about studies published) • Data Source History <ul style="list-style-type: none"> • Age, updates and accuracy • Historical data availability • Ethics (specific protocols, duration of approval process, restrictions) • Representativeness <ul style="list-style-type: none"> • Age–sex profiles • Ethnicity information • Deprivation indices • Additional consents 		<p>Data Interoperability Level</p> <ul style="list-style-type: none"> • Unique identifiers and pseudonymisation • Links to primary care data, national health identifiers <p>Data Types</p> <ul style="list-style-type: none"> • Coding systems (ICD, ICPC, SNOWMED, CTV3/Read) • Mental Health nursing diagnosis • Classification systems (procedures, genetic information etc.) • Episodes and free-text input • Guidance between diagnosis and therapy <p>Metadata</p> <ul style="list-style-type: none"> • Structure • Association of data with clinicians and practices • Data quality 	<p>Record System</p> <ul style="list-style-type: none"> • Recognisable formats • Archetypes • Extraction API • EPR interface • Hardware <p>Software</p> <ul style="list-style-type: none"> • Network connection • Local/remote query execution • Efficiency and backups <p>Team and Users</p> <ul style="list-style-type: none"> • Collection process • Training • Team size <p>Data Modelling and Size</p> <ul style="list-style-type: none"> • Extract type • Statistical analysis • Audit-trails • Patient number and recordings 	<p>Organisational Structure</p> <ul style="list-style-type: none"> • State financed health system • Private insurance • GP referrals • Multi-practice registration • Emergency treatment • Childcare, maternity, dental mental health and sexual health as part of primary health care • National associations for primary care practitioners 	<p>Social and Cultural restrictions:</p> <ul style="list-style-type: none"> • Legal, ethical, social, economical, political, religious factors <p>Other factors:</p> <ul style="list-style-type: none"> • Factors able to affect data validity <p>Languages:</p> <ul style="list-style-type: none"> • Primary and Secondary language being used • English as accepted language for ethical bodies 	<p>Data quality & improvement relevant to use-cases</p> <ul style="list-style-type: none"> • Plans for improving the data (extending coverage, improving completeness and reducing bias) <p>Demographics and other data</p> <ul style="list-style-type: none"> • Age, gender, ethnicity, country of birth • Physical examinations (e.g. weight, blood pressure, etc) • Medical history (e.g. smoking habits, activity, family history, etc) • Access to laboratory and imaging results

Figure 1 Detail of the questionnaire schema. At the data-source level, A is completed by genetic databases only

(CMR) and other data repositories,¹⁹ audit trails and the size of the database. Finally, the third section, the macro level, explored issues relating to the nature of the health system, sociocultural factors and issues relating to funding, the purpose and restrictions on the use of the data. Study-specific questions make up the final part of the survey instrument (Figure 1).²⁰

Piloting the TIRRE questionnaire with the first ten repositories

We initially contacted 10 sites who were interested in participating in the TRANSFoRM project by email, if available, and followed this up with telephone calls. We collected data through a structured telephone interview. These 10 repositories were based throughout Europe. These first 10 sites were also stakeholders in our requirements analysis.¹⁸

The pilot sample

The pilot sample was comprised of mainly primary care databases (50%, $n = 5$); this was principally because we had better links to these databases and this made conducting the telephone pilot easier. The sample included epidemiological cohorts (20%, $n = 2$), disease registries (20%, $n = 2$) and a genomic database (10%, $n = 1$). All of the organisations questioned held data for the purpose of research. We set out to recruit a representative sample.

Development of the online questionnaire using skip logic

We then developed and introduced an online version of our survey instrument whilst continuing to offer the telephone method in parallel for repositories that preferred to pass on information this way. The survey was comprehensive and consisted of 160 questions. We used a software package from SurveyMonkeyTM to create and develop the online questionnaire. To avoid

Table 2 Categories of data collection and min-to-max number of questions depending on the extent to which skip logic is utilised

Level	Category	Primary care data	Genetic database	Cancer registry	Others (social care data, cohorts)	Number of questions
Micro	Data source	30–46	36–54	30–46	33–51	54
	Data interoperability	30–43	31–43	31–43	31–43	43
	Subtotal	60–89	67–97	61–89	64–94	97
Meso	Record system	5–30	5–30	5–30	5–30	30
Macro	Organisational	15–15	15–15	15–15	15–15	15
	Sociocultural	6–10	6–10	6–10	6–10	10
	Subtotal	21–25	21–25	21–25	21–25	25
Study	Use-case specific	5–8	5–8	5–8	5–8	8
Total		Min: 91; Max: 152	Min: 98; Max: 160	Min: 92; Max: 152	Min: 95; Max: 157	160

unnecessary repetition in data collection and to ensure that only relevant information for each type of repository was collected, the online version was designed using skip logic. Skip logic also has the benefit of making the survey instrument more streamline, reducing the number of questions each respondent is required to answer (Table 2).

Ethical considerations

There was no formal ethics board review. This survey only seeks to report information about the capacity and capability of information sources to be combined to conduct research studies. The survey does include a self-declaration that the data held were subject to an ethical process. We did not consider any issues of consent or privacy, beyond exploring a site's track record for studies and that there was a mechanism for obtaining consent for any specific study.

Results

The pilot sample

Seven of the ten sites confirmed that there were no restrictions on the use of their data for ethically approved research studies and all had had studies completed and published using their data source.

Seven were able to confirm that their data sets could be used for cross-sectional studies, three had the ability to link to other genetic databases and three had access to other methods of data collection such as social security systems. Five of the repositories required additional patient consent to be completed to use their data for research; of these, four confirmed that individual consent was not required if strong identifiers were removed. The same five stated that individual consent was required to: (1) link single nucleotide polymorphisms (SNPs) to phenotype or primary care data studies, and (2) identify individuals participating in cohort, cross-sectional studies or for recruitment into trials. Eight repositories were able to confirm that there was a local or network-specific protocol for ethical approval that needed to be followed for their collections, and seven said that there were restrictions on the use of their data for ethically approved research studies. Seven of the repositories confirmed their data source did not employ a unique identifier to allow linkage between primary care, hospital and national health data. The majority held historical data (e.g. from a patient's birth) and all except one could provide age-sex profile data from their source.

The sample was drawn as follows: there were five primary care sites, one from each of France, Germany, Denmark, Italy and Finland. France also provided two epidemiological cohorts; there was a genomics database from the UK and a disease registry from Switzerland completed the sample.

Feedback and developing the questionnaire

The initial telephone interview took around 1.5 hours to complete. This was deemed too long, felt onerous for the person supplying the information and provided no obvious benefit to them. With this in mind, we sought to improve the questionnaire by three methods. First, we analysed the 104 questions that had a yes/no response and explored whether they are more likely to be positive in a database that is eligible to complete the use-case. Second, we looked at the multiple response questions, simplifying them into single response questions, we looked for common categories that allowed us to reduce the number of free-text responses and investigated the skip logic sections. Third, we removed questions that did not discriminate between use-case eligible or not. The study team then reviewed questions for potential removal from the TIRRE survey instrument, as some non-discriminatory data may still be useful and required (e.g. all databases need to be able to state the size of their population). As a result of the above activities, we have been able to remove the questions that we do not think are essential, reducing the number to 139.

For the second version of the TIRRE survey instrument we removed ‘Don’t know’ as an answer option in order to prompt a yes or no response. Many questions from the first version of the TIRRE survey were combined and rewritten for the second version in order to simplify answer options.

To improve the identification of data sources with greater potential for generating linkages we added a new Question 4 ‘Does the data source provide national coverage?’, and to assess the readiness of a data source to participate in a specific study we added new questions ‘From first receiving a formal request for study data, how long do you anticipate it would take you to export the data to researchers?’ (Question 136 in TIRRE2) and ‘In the last 12 months what is the

mean time taken from receiving any formal request for study data through to exporting that data to researchers?’ (Question 137 in TIRRE2). We removed Questions 33–36 (asking for text data about the actual publications) as we felt this data could be asked for or checked subsequently if necessary. In order to simplify the responses we rewrote many questions. For example, we changed Question 38 (the original Question 42 in TIRRE1) from ‘How long does it typically take to get ethical approval?’ to ‘Does it typically take less than 6 months to get ethical approval’. This process is subject to on-going development.

Summary results

We created a screening tool to assess the readiness of clinical data repositories around the world to participate in linked research. By utilising skip logic we can streamline the TIRRE tool, making it more specific to each type of stakeholder and reducing the number of questions needed (Tables 3 and 4). Questions selected for the final online version were judged to be essential to ascertain key information about the suitability of data repositories from at least two members of the same group.^{21,22}

Sensitivity analysis

Eight of the ten data-providing stakeholders could quote peer review publications produced from their data and indexed in PubMed.²³ Four estimated that between 30 and 100 peer review articles had been published in the last five years; of the remaining six, two estimated that between 11 and 20 articles had been published, one estimated that 2 to 5 articles had been published, and one stated that no publications had been generated based on their data.

Table 3 Summary of issues for data repositories taking part in the initial pilot

Project stakeholder	Specific information needs	Best source of information needed	Planned method of delivery	Timing considerations
Data providers	1. Return on investment of time – studies, funding 2. Strategic interest	Who will get a funded study? Track record of researcher	Engage once benefits	Have standard processable outputs readily available

Table 4 Categories of data collection and min-to-max number of questions after application of the skip logic

Analysis by respondent	Number of questions								
	Should	Primary care		Cancer registry		Genetic		Overall	
		Mean	Response	Mean	Response	Mean	Response	Mean	Response
Categorical questions	<i>n</i>	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
Data source	35	22.41	64	24.00	69	25.00	71	23.39	67
Data inter-operability	31	29.34	95	28.80	93	26.25	85	28.54	92
Micro level	66	51.76	78	52.80	80	51.25	78	51.93	79
Meso level/ record system	12	3.10	26	2.33	19	2.67	22	2.80	23
Organisational	15	15.00	100	15.00	100	15.0	92	14.73	98
Socio-cultural	6	5.17	86	5.47	91	4.50	75	5.11	85
Macro level	21	20.17	96	20.47	97	18.25	87	19.84	94
Overall	104	79.90	76.82	80.60	77.50	83.27	80.07	80.76	77.66

Discussion

Principal findings

The TIRRE instrument is a screening tool to assess both the potential of a database to take part in use-cases and the feasibility of conducting linked research; although we recognise that further testing may be required before a data source participates in a study. The TIRRE survey systematically collects the extensive amount of information required to assess readiness to participate in studies linking primary care and either genetic databases or cancer registry data. It is based on a requirements analysis. Although we had difficulty in getting databases to complete the questionnaire, when we did get a response the completeness of the information gathered was high and proved useful in identifying their potential to participate in linked research. There are currently no other international sites available to enable brokerage between databases willing to participate in research.

Implications of the findings

The TIRRE survey is the best first step towards assessing the 'linkability' of databases in one locality and the extent to which a study might be run across several localities. It has the potential to measure readiness in several dimensions:

- data quality readiness from micro-level information
- record readiness and the technical ability of the computer systems that hold the data to export it reliably and
- governance readiness, including any information about sociocultural barriers to utilisation.

Collectively, these dimensions inform about the readiness to scale the study beyond a single database – which we term scalability readiness. However, any initial screening process may need to be followed up by a detailed assessment of whether the data set needed for a given study can be elicited from the data repositories.

Comparison with the literature

It is possible to draw comparisons between the complexity of this task and the existing successful projects that involve linking data. However, the successful data repositories in the UK have all been based on a single brand of general practice electronic health record (EHR) system. The General Practice Research Database (GPRD), recently renamed the Clinical Practice Research Datalink (CPRD), extracts data from a brand called In-Practice Systems (INPS) Vision, although it is set to expand its coverage. The Health Improvement Network (THIN) also collects data from INPS.²⁴ Q-Research takes data from the EMIS EHR²⁵ and other sentinel networks all follow the same pattern. Even the relatively simple task of linking data from this small number of brands of computer within the UK has proved challenging, in terms of both creating a summary care record²⁶ and developing a common data-extraction system.²⁷ Overall, we are unable to find anything similar to TRANSFoRm in academia. Wider use of ontologies may provide a framework for better information sharing.^{28,29}

Little has been written about the concept of research readiness and its importance for studies requiring data from more than one source. Early conceptual models for primary care research networks focused on the structures, process and outcomes – without thinking of readiness as a key concept.³⁰ Research readiness has been introduced in the UK at the individual practice level; although this is through a system of self-accreditation.³¹ Internationally, the Electronic Health Record for Clinical Research (EHRCR) Project has produced a functionality profile which outlines the high-level requirements necessary for EHR systems to be considered as a reliable data source in line with the appropriate regulations governing clinical research. This profile will be used as the basis for the certification of EHR systems used in clinical research with the aim of increasing the level of comfort of practitioners, the research community and regulators with the practice of storing source data in EHR systems;³² although this approach does not appear to have been adopted widely.

Limitations of the method

An initial survey, provided from a previous work package, provided limited information on the suitability and readiness of databases.¹⁷ Redefinition of the variables required for these databases enabled us to conduct a further survey. There was no real incentive for data repositories to supply us with the data required, as there was not a reciprocal offer of benefit. Successful completion of the survey was a lengthy process, taking in excess of 1 hour. As a consequence,

our results inevitably underestimate the number of sites where this type of research can be conducted.

Inevitably, simplification needed to construct effective middleware. The use-cases have developed appropriate but complex models. The provenance of data defined within the work package has been broad, seeking to map the provenance of data from the original observation, but going beyond the scope of the study use-cases.

Finally, our approach was framed by the English context,³³ and communication in English. However, in mitigation we also have a European perspective on health system data.³⁴

Call for further research

We need to conduct test–retest studies to assess the reliability of the survey instrument. We should conduct simulated and real studies with data extractions to test its validity.

Conclusions

A large complex set of data is needed to know if it will be possible to link primary care and either a disease registry or genetic database. This set of data can either be classified by the level of granularity or as a business or data requirement. The TIRRE instrument is evolving into a tool that can be used to assess general suitability and readiness to participate in linked research studies.

ACKNOWLEDGEMENTS

The authors thank Paul van Royen and Robert Verheij for their comments on the manuscript; IMIA and EFMI for supporting the primary care informatics working groups. Antonis Ntasioudis for his input into TIRRE. TRANSFoRm is supported by the European Commission – DG INFSO (FP7 2477).

REFERENCES

- 1 European Union, Europe's Information Society. *Communication from the Commission to the Council, the European Parliament, the European Economic and Social Committee and the Committee of the Regions. e-Health – making healthcare better for European citizens: an action plan for a European e-Health Area* {SEC(2004)539}, 2004. www.eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:52004DC0356:EN:NOT
- 2 TRANSFoRm (a Translational Research and Patients safety In Europe). www.transformproject.eu

- 3 [International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature* 2001;409\(6822\):860–921.](#)
- 4 [International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature* 2004;431\(7011\):931–45.](#)
- 5 [Hirtzlin I, Dubreuil C, Préaubert N *et al.* An empirical survey on biobanking of human genetic material and data in six EU countries. *European Journal of Human Genetics* 2003;11\(6\):475–88.](#)
- 6 [McEwen J. Forensic DNA data banking by state crime laboratories. *American Journal of Human Genetics* 1995; 56\(6\):1487–92.](#)
- 7 [Schneider B. Progress in cancer control through cancer registries. *CA: A Cancer Journal for Clinicians* 1958; 8\(6\):207–10.](#)
- 8 [Navarro C, Martos C, Ardanaz E *et al.* Population-based cancer registries in Spain and their role in cancer control. *Annals of Oncology* 2010;21\(Suppl 3\):iii3–13.](#)
- 9 [Gatta G, Capocaccia R, Trama A, Martínez-García C and the RARECARE Working Group. The burden of rare cancers in Europe. *Advances in Experimental Medicine and Biology* 2010;686:285–303.](#)
- 10 [Francisci S, Capocaccia R, Grande E *et al.* The cure of cancer: a European perspective. *European Journal of Cancer* 2009;45\(6\):1067–79.](#)
- 11 [de Lusignan S and Chan T. The development of primary care information technology in the United Kingdom. *Journal of Ambulatory Care Management* 2008;31\(3\): 201–10.](#)
- 12 [Schade CP, Sullivan FM, de Lusignan S and Madeley J. e-Prescribing, efficiency, quality: lessons from the computerization of UK family practice. *Journal of the American Medical Informatics Association* 2006;13\(5\): 470–5.](#)
- 13 [de Lusignan S and van Weel C. The use of routinely collected computer data for research in primary care: opportunities and challenges. *Family Practice* 2006; 23\(2\):253–63.](#)
- 14 [de Lusignan S, Liaw S-T, Krause P *et al.* Key concepts to assess the readiness of data for international research: data quality, lineage and provenance, extraction and processing errors, traceability, and curation. *Yearbook of Medical Informatics* 2011;6\(1\):112–20.](#)
- 15 [de Lusignan S, Krause P, Michalakidis G *et al.* Business process modelling is an essential part of a requirements analysis for studies linking heterogeneous clinical databases. *Yearbook of Medical Informatics* 2012;7\(1\): 34–43.](#)
- 16 [de Lusignan S, Metsemakers JF, Houwink P, Gunnarsdottir V and van der Lei J. Routinely collected general practice data: goldmines for research? A report of the European Federation for Medical Informatics Primary Care Informatics Working Group \(EFMI PCIWG\) from MIE2006, Maastricht, The Netherlands. *Informatics in Primary Care* 2006;14\(3\):203–9.](#)
- 17 [Dolin RH and Alschuler L. Approaching semantic interoperability in Health Level Seven. *Journal of the American Medical Informatics Association* 2011;18\(1\): 99–103.](#)
- 18 [Michalakidis G, Kumarapeli P, Ring A, van Vlymen J, Krause P and de Lusignan S. A system for solution-orientated reporting of errors associated with the extraction of routinely collected clinical data for research. *Studies in Health Technology and Informatics* 2010; 160:724–8.](#)
- 19 [Santos M, Bax M and Kalra D. Building a logical EHR architecture based on ISO 13606 standard and semantic web technologies. *Studies in Health Technology and Informatics* 2010;160:161–5.](#)
- 20 [Clinical Informatics. TRANSFoRM International Research Readiness \(TIRRE\) instrument. \[www.clininf.eu/jointirre/info.html\]\(http://www.clininf.eu/jointirre/info.html\)](#)
- 21 [de Lusignan S, Peace C, Shaw N *et al.* What are the barriers to conducting international research using routinely collected primary care data? *Studies in Health Technology and Informatics* 2011;165:135–40.](#)
- 22 [de Lusignan S, Liaw S-T, Krause P *et al.* Key concepts to assess the readiness of data for international research: data quality, lineage and provenance, extraction and processing errors, traceability, and curation. *Yearbook of Medical Informatics* 2011;6\(1\):112–20.](#)
- 23 [US National Library of Medicine, National Institute of Health. *PubMed Medline*. \[www.ncbi.nlm.nih.gov/pubmed\]\(http://www.ncbi.nlm.nih.gov/pubmed\)](#)
- 24 [Blak BT, Thompson M, Dattani H and Bourke A. Generalisability of the Health Improvement Network \(THIN\) database: demographics, chronic disease prevalence and mortality rates. *Informatics in Primary Care* 2011;19\(4\):251–5.](#)
- 25 [Hippisley-Cox J, Stables D and Pringle M. QRESEARCH: a new general practice database for research. *Informatics in Primary Care* 2004;12\(1\):49–50.](#)
- 26 [Greenhalgh T, Stramer K, Bratan T, Byrne E, Russell J and Potts HW. Adoption and non-adoption of a shared electronic summary record in England: a mixed-method case study. *BMJ* 2010 Jun 16;340:c3111.](#)
- 27 [NHS Information Centre. *GP Extraction Service*. \[www.ic.nhs.uk/gpes\]\(http://www.ic.nhs.uk/gpes\)](#)
- 28 [de Lusignan S, Liaw ST, Michalakidis G and Jones S. Defining datasets and creating data dictionaries for quality improvement and research in chronic disease using routinely collected data: an ontology-driven approach. *Informatics in Primary Care* 2011;19\(3\):127–34.](#)
- 29 [Liaw ST, Rahimi A, Ray P *et al.* Towards an ontology for data quality in integrated chronic disease management: a realist review of the literature. *International Journal of Medical Informatics* 2012 Nov 1. doi:pii: S1386–5056\(12\)00193–1. 10.1016/j.ijmedinf.2012.10.001.](#)
- 30 [Clement S, Pickering A, Rowlands G, Thiru K, Candy B and de Lusignan S. Towards a conceptual framework for evaluating primary care research networks. *British Journal of General Practice* 2000;50\(457\):651–2.](#)
- 31 [Royal College of General Practitioners. *Research Ready Self-accreditation*. \[www.rcgp.org.uk/clinical_and_research/circ/research_knowledge_transfer/research_ready.aspx\]\(http://www.rcgp.org.uk/clinical_and_research/circ/research_knowledge_transfer/research_ready.aspx\)](#)
- 32 [EHRCR \(Electronic Health Records for Clinical Research\) Working Group Sponsored by eClinical Forum and PhRMA EDC/eSource Task Group. *Practical considerations for clinical trial sites using electronic health records \(EHRs\) in support of clinical research addressing regulatory considerations*, 2009. \[www.clinpage.com/images/photos/EHRCR_Apr09.pdf\]\(http://www.clinpage.com/images/photos/EHRCR_Apr09.pdf\)](#)

- 33 [Swindells M and de Lusignan S. Lessons from the English National Programme for IT about structure, process and utility. *Studies in Health Technology and Informatics* 2012;174:17–22.](#)
- 34 [de Lusignan S, Minmogh C, Kennedy J, Zeimet M, Bommezijn H and Bryant J. A survey to identify the clinical coding and classification systems currently in use across Europe. *Studies in Health Technology and Informatics* 2001;84:86–9.](#)

ADDRESS FOR CORRESPONDENCE

Professor Simon de Lusignan
Department of Health Care Management and Policy
University of Surrey
Guildford
GU2 7XH
UK
Tel: +44 1483 863089
Email: s.lusignan@surrey.ac.uk

Accepted February 2013