Research article

A systematic scoping review of the domains and innovations in secondary uses of digitised health-related data

Author address for correspondence
Ann R R Robertson
Centre of Medical Informatics, Usher Institute of Population Health Sciences and Informatics,
Teviot Place, The University of Edinburgh
Edinburgh, UK
Email: A.R.R.Robertson@ed.ac.uk

Ann R R Robertson
Centre of Medical Informatics, Usher Institute of Population Health Sciences and Informatics,
The University of Edinburgh, Edinburgh, UK

Ulugbek Nurmatov
Centre of Medical Informatics, Usher Institute of Population Health Sciences and Informatics,
The University of Edinburgh, Edinburgh, UK

Harpreet S Sood
National Health Service (NHS) England, London, UK

Kathrin Cresswell
Centre of Medical Informatics, Usher Institute of Population Health Sciences and Informatics,
The University of Edinburgh, Edinburgh, UK

Pam Smith
Nursing Studies, School of Health in Social Science, The University of Edinburgh, Edinburgh, UK

Aziz Sheikh
Centre of Medical Informatics, Usher Institute of Population Health Sciences and Informatics,
The University of Edinburgh, Edinburgh, UK

ABSTRACT

Background: Substantial investments are being made in health information technology (HIT) based on assumptions that these systems will save costs through increased quality, safety and efficiency of care provision. Whilst short-term benefits have often proven difficult to demonstrate, there is increasing interest in achieving benefits in the medium and long term through secondary uses of HIT-derived data.

Aims: We aimed to describe the range of secondary uses of HIT-derived data in the international literature and identify innovative developments of particular relevance to UK policymakers and managers.

Methods: We searched nine electronic databases to conduct a systematic scoping review of the international literature and augmented this by consulting a range of experts in the field.

Results: Reviewers independently screened 16,806 titles, resulting in 583 eligible studies for inclusion. Thematic organisation of reported secondary uses was validated during expert consultation (*n* = 23). A primary division was made between patient-identifiable data and datasets in which individuals were not identified. Secondary uses were then categorised under four domain headings of: i) research; ii) quality and safety of care provision; iii) financial management; and iv) healthcare professional education. We found that innovative developments were most evident in research where, in particular, dataset linkage studies offered important opportunities for exploitation.

Conclusions: Distinguishing patient-identifiable data from aggregated, de-identified datasets gives greater conceptual clarity in secondary uses of HIT-derived data. Secondary uses research has substantial potential for realising future benefits through generating new medical knowledge from dataset linkage studies, developing precision medicine and enabling cross-sectoral, evidence-based policymaking to benefit population-level well-being.

Keywords: Medical informatics, health services research, systematic scoping review

BACKGROUND

Healthcare still trails behind other major industries in fully exploiting information and communication technologies (ICTs) to maximise quality, safety and efficiency in service delivery. It can be argued that this is partly due to the complexity of organising and delivering healthcare services and to the challenges of introducing standardised ICT systems across healthcare settings where these are diverse and largely autonomous organisations. Yet considerable effort and capital investments are being made in the United Kingdom (UK) national health services and in healthcare organisations internationally to procure and implement ICT systems, also known as eHealth or health information technology (HIT).[1,2] Such investments have been justified by assumptions that routine use of HIT should lead to improved patient outcomes and to cost saving efficiencies in service delivery, for example by streamlining care processes.[3,4] Recent work, however, suggests that in practice benefits from HIT can be hard to identify, at least in the short term. It has, for instance, been found that some processes can become more time consuming for some staff during the early years of using a new HIT system.[3,4]

Such disappointing evidence for the anticipated quick gains and returns on investments in HIT could potentially jeopardise continued spending on HIT initiatives. Unrealistic assumptions about the timelines for delivering benefits, for example, from core systems such as electronic health records and ePrescribing systems, place an emphasis on early measurable gains, whereas more significant advantages to healthcare and society might accrue in the medium to long term and then particularly through innovative uses of the wealth of health-related digital data that become available.[5–9] Secondary uses of data – the use and re-use of clinical and administrative data other than for the direct clinical care of specific patients – may present the greater opportunity for realising benefits from HIT investments, with such benefits emerging more slowly.

In 2007, the American Medical Informatics Association (AMIA) identified the then current areas of secondary uses of health-related digital data in USA settings.[10] This systematic scoping review, part of a larger, mixed-methods investigation into maximising the safe and secure exploitation of HIT-derived data in the UK context, aimed to build on that earlier USA work in order to provide an updated, international framework of secondary uses. Our focus was on current and potential future developments of particular relevance for UK policymakers and health service managers.

METHODS

We conducted a systematic scoping review. According to Arksey and O'Malley,[11] a scoping study is a type of literature review that can serve to 'map' a field of interest; unlike a systematic literature review, it is unlikely to address a narrowly defined research question or to assess the quality of included studies. This approach is well suited to exploring under-researched or emerging fields of study, where empirical evidence is limited. Our systematic scoping review was guided by the six-stage methodological framework developed by Levac et al.[12] from the Arksey and O'Malley's original framework (Box 1).

> **Box 1 Summary of methodological guidelines for systematic scoping reviews, from Levac et al.[11]**
>
> 1. Identify the research question/purpose of the scoping review
> 2. Identify relevant studies
> 3. Select studies
> 4. Chart the data
> 5. Collate and summarise findings
> 6. Consult stakeholders

Identifying relevant studies

With assistance from an experienced medical librarian, we developed, tested and modified a strategy to search the published literature from the beginning of 2000 to the end of 2013 in nine electronic databases. We searched the following databases: MEDLINE, EMBASE, The Cochrane Database of Systematic Reviews, Database of Abstracts of Reviews of Effects, The Cochrane Central Register of Controlled Trials, The Cochrane Methodology Register, The Health Technology Assessment Database, National Health Service (NHS) Economic Evaluation Database, and The National Research Register, as well as in LILACS, Lexis Library, and Google Scholar for grey literature. The terms in the first set of search terms related to HIT systems and were based on a literature search strategy previously employed in a systematic review of the eHealth literature.[3] Terms in the second search set related to secondary uses of healthcare data, drawing on the 2007 taxonomy of secondary uses developed by AMIA.[10] (Appendix 1). Terms within groups were combined using the Boolean operator "OR" and the groups combined with the Boolean operator "AND" (Appendix 2). We applied no language or publication status restrictions.

Study selection

After initial screening by the team and removing duplicate publications, three researchers (AR, UN and KC) independently checked titles and abstracts against the inclusion/exclusion criteria. We included empirical studies that reported information about secondary uses of data held in core HIT systems in developed countries. Publications were excluded if they fell outside the scope of interest, for example, those reporting on technologies not associated with core HIT functionalities (for instance reports of speech recognition functionality). We excluded studies reporting on HIT implementations in developing countries because of the contextual differences between healthcare and its delivery in developing and developed countries. We then retrieved and reviewed the full texts of potentially eligible publications.

Charting the data

We used customised Excel forms to extract data from each of the full text papers eligible to be included in our review. The variables that were recorded by three researchers (AR, HS and UN) were author and year, title of the study, country of origin, keywords and the area of secondary uses reported, and if a study was deemed by the reviewer to offer an example of a new development in secondary uses.

Collating and summarising the extracted data

We used a thematic, qualitative content analysis approach[12,13] to organise the various areas of secondary uses identified from the review into broad domains of secondary uses, resolving any uncertainties by discussion among the researchers who had charted the data and within the wider research team.

Consultation with experts

We discussed our preliminary findings with a range of national and international experts to seek validation of our thematic organisation into domains of secondary uses and any additional insights into innovative developments in secondary uses. These individuals were selected based on their involvement in activities related to using data held in HIT systems in the UK, with additional experts beyond the UK being invited from regions with an international reputation for current work in this field. Consultees included policymakers, health professionals, academics (including researchers) and representatives of the pharmaceutical industry, the legal profession and the third sector. We approached 28 potential consultees (declined = 1; no response or subsequently could not be contacted = 4), leading to 23 interviews with participants throughout the UK and in Australia, Canada and the USA. One participant subsequently withdrew consent, reporting new workplace restrictions on giving interviews and consequently that audio file and transcript were deleted from our dataset. During the consultation stage of the scoping review, consultees were asked to highlight areas illustrative of developments in secondary uses research.

Ethical approval

Ethical approval was not required for the systematic scoping review. The consultation phase was conducted within a related interview study that formed part of the larger, mixed-methods project from which we are reporting the scoping review component here. We obtained ethical approval for the interview study from The University of Edinburgh, and each participant gave informed consent prior to taking part.

RESULTS

Our search strategy identified 20,551 potentially relevant papers. After deduplication, 16,806 papers were included for initial screening; a further 15,089 papers were excluded because they did not meet the inclusion criteria. 1717 retained abstracts were reviewed, of which 1134 papers were defined as background papers (for example papers describing HIT infrastructure), resulting in 583 studies being included in the review. The results are presented as a Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) study flow diagram, shown in Figure 1.

Source of studies

The included studies represented a spread of developed countries, with the most prominent being the USA. They also included Canada, Australia, the UK, other European countries, Scandinavian nations and The Netherlands (Figure 2).

Secondary uses

The publications referenced a range of areas in which health-related digital data were being used beyond supporting individuals' clinical care. The scope of reported secondary uses included for conducting epidemiological and pharmacovigilance research studies, for facilitating recruitment to randomised controlled trials[14–16] and for carrying out audits and benchmarking studies.[17,18] We also found secondary uses being used for financial and service planning, incident tracking, the teaching of clinical staff and billing.[19,20]

The examples of secondary uses came from within a single healthcare organisation (for example local audits), across healthcare settings (such as in service planning) and from dataset linkage studies.

Dataset linkage research

Innovative developments were most evident in the research domain, with ongoing efforts in several developed countries to establish the research infrastructure for dataset linkage studies.[21–26] For instance, researchers were able to use Scotland-wide routinely collected hospital admission data combined with death certificate data to show that legislation to ban smoking in public places in Scotland was followed by a reduction in hospital admissions for childhood asthma.[27] In Denmark, researchers also used health dataset linkage to conduct a nationwide seven-year study of everyone aged 18–36, using national registries, death certificates and primary care data to investigate the relative and absolute risks of sudden cardiac death in young Danes with a prior myocardial infarction.[28]
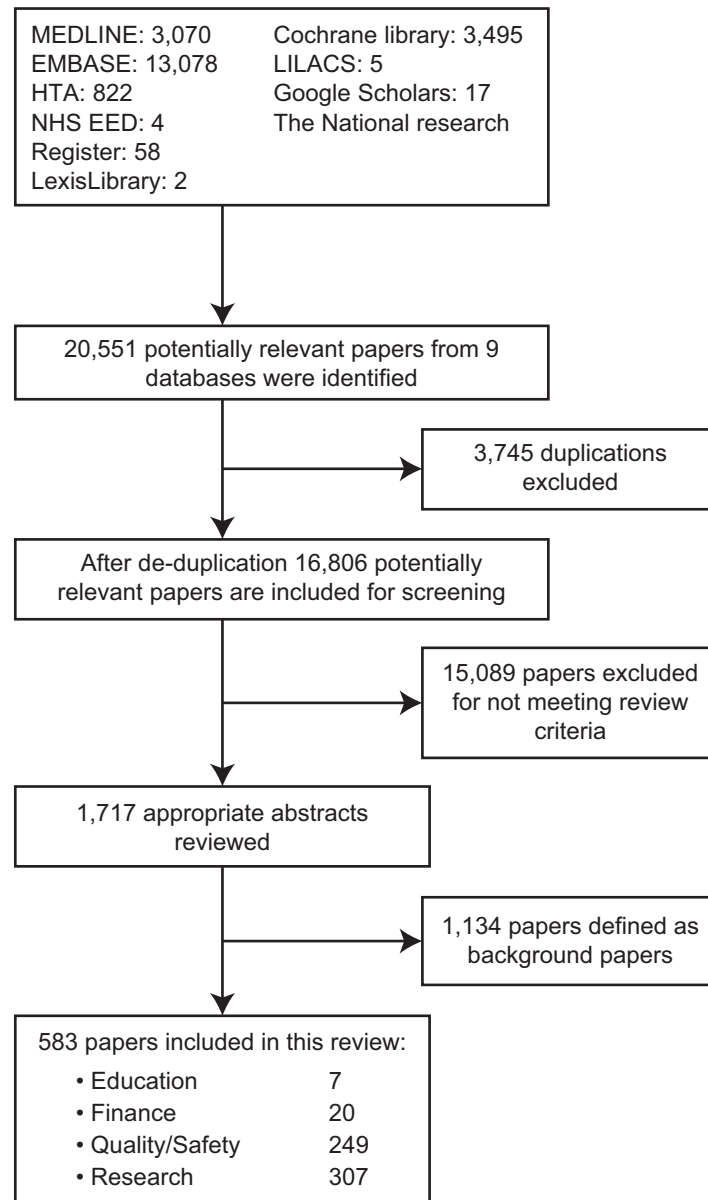
MEDLINE: 3,070 Cochrane library: 3,495
EMBASE: 13,078 LILACS: 5
HTA: 822 Google Scholars: 17
NHS EED: 4 The National research
Register: 58
LexisLibrary: 2

20,551 potentially relevant papers from 9 databases were identified

3,745 duplications excluded

After de-duplication 16,806 potentially relevant papers are included for screening

15,089 papers excluded for not meeting review criteria

1,717 appropriate abstracts reviewed

1,134 papers defined as background papers

583 papers included in this review:
- Education 7
- Finance 20
- Quality/Safety 249
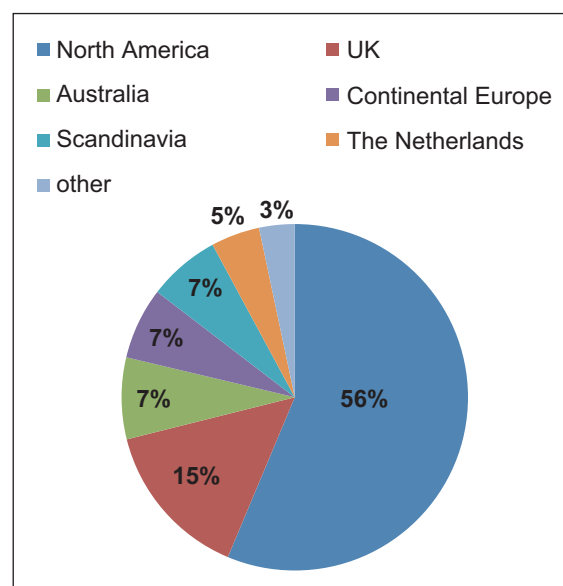- Research 307

Figure 1 PRISMA flow diagram

Figure 2 Countries of origin of the papers included in the scoping review

In addition to data linkage studies using health-related datasets, existing examples of studies using cross-sectoral data linkage included the linking of population-wide health and justice datasets in Western Australia to study hospitalisations among exprisoners during the first year after their release[29] and seeking evidence to help plan healthy neighbourhoods across the lifespan by investigating measures of the built environment linked to health outcomes and to self-reported health behaviours.[30]

Domains of secondary uses

Overall, the majority of examples of secondary uses identified from scoping the literature could be categorised under the broad heading of research. Research was followed by a second large domain of quality and safety of care provision, which included audit. We grouped all of the studies in the review thematically into a total of four broad domains of current secondary uses: 1) research ($n = 307$); 2) quality and safety of care ($n = 249$); 3) financial management ($n = 20$); and 4) education ($n = 7$) (Figure 3). An initial long list of secondary uses generated from the scoping review, from which the four domains were derived, is shown in Appendix 3.

Consultation phase of scoping review

We approached 28 people (declined = 1; no response or subsequently could not be contacted = 4), leading to 23 expert participants throughout the UK and in Australia, Canada and the USA. During this stage of the scoping review, consultees were asked to highlight areas illustrative of new developments in secondary uses research. They drew attention to investigations into risk factors, treatments and disease outcomes (notably, in Scotland, diabetes-related studies), drug safety, and policy evaluation, service delivery and public health.[31–34]

Consultees were asked to comment on the thematic organisation of secondary uses identified from scoping the literature into four domains. In addition to listing the four current domains, it was suggested it would be helpful first to distinguish between secondary uses involving data containing identifiers for patients – essential for providing direct, clinical care and also for some secondary uses, for instance for tracing individuals affected by contaminated surgical instruments in crisis management – and aggregated, deidentified datasets (where deidentified data were also variously known as anonymised and pseudonymised data). Keeping that distinction to the fore was considered important for policymakers and managers aiming to maximise HIT-derived benefits because of the potential for significant new research findings that were dependent on exploiting large quantities of deidentified aggregated data. Confusing data with and without patient identifiers could negatively impact on public support for secondary uses. Patients' privacy, confidentiality and consent for the use and reuse of data where those data identified individual patients were recognised as important concerns to many people.

Looking towards the future of HIT-derived data and secondary uses, consultees spoke of expanding the range of health-related datasets that were available to researchers in the UK to include general practice, imaging, genomic and biotech data, and datasets from non-health sectors such as education, housing and justice. In the UK, the potential for the Farr Institute[35] to be working in close collaboration with the Administrative Data Research Network (ADRN)[36] was highlighted as a positive step for developing cross-sectoral research work. It was believed that population well-being should benefit from cross-sectoral dataset linkage research with such studies generating an evidence base to underpin UK policy decisions and policy evaluations beyond specifically health policy. However, this could not be achieved without national workforce planning and training in order to have sufficient staff with the necessary range of technical and methodological skills to work in data linkage.

In addition to envisaged developments in – and longer-term benefits from – dataset linkage studies, progress in natural language processing software should increasingly allow researchers to take advantage of uncoded text in electronic health records. Those data and patient reported measures both would add to the digital data likely to become more widely available for secondary use research in the future.

DISCUSSION

We searched nine international electronic databases, screened 16,806 titles and found 583 eligible studies. The systematic scoping review identified secondary uses of digitised health-related data in the domains of research (the largest category), quality and safety of service provision, financial management, and education. Innovations in secondary uses were most evident in the research domain with the development of dataset linkage studies. Consultation with experts confirmed that research linking datasets – both linking health datasets with each other and linking between health and datasets from other sectors – would in their opinion continue to expand and to deliver health-related and wider societal benefits from investments in HIT systems.
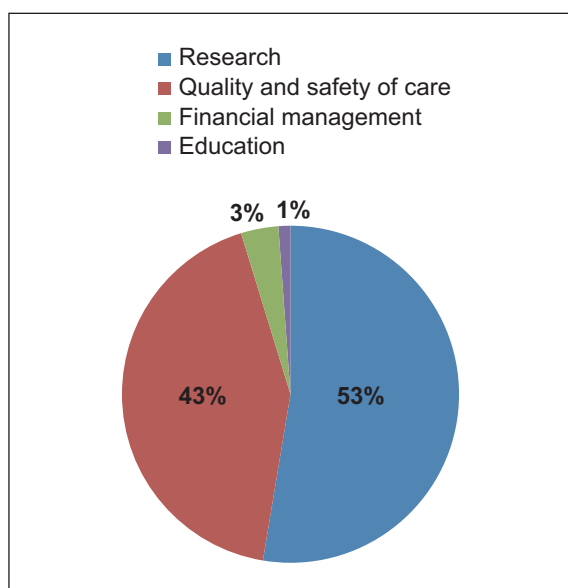
Figure 3 Domains of secondary uses identified in the scoping review

- Research
- Quality and safety of care
- Financial management
- Education

3% 1%
43% 53%

This is the first UK-focused systematic scoping review of secondary uses, updating the previous work in this area undertaken elsewhere.[10] The publicly funded NHS in the UK and the availability of national and regional datasets contribute to a UK-specific context for secondary uses of health-related digital data and likely offer particularly strong potential for innovative research that exploits dataset linkages.

While the UK context and a growing emphasis on dataset linkage studies are quite distinctive, the range of areas of secondary uses identified in our literature review is similar to the areas of secondary uses previously identified despite the passing of time since that earlier work from the USA.[10] Domains for secondary uses of health-related digital data may have reached a level of stability, at least for the foreseeable future. The more dynamic aspects appear likely to be contextual factors, for example national and international legislation controlling personal data, and further developments within a given secondary use domain, such as within the research domain. Understanding where the most potential for developing secondary uses currently lies and appreciating the importance of distinguishing clearly between data that identify patients and data that are aggregated and deidentified are a resource for UK policymakers who are developing plans and policies related to secondary uses of health-related digital data and for all those aiming to maximise returns from investing in HIT systems.

Strengths and limitations

The main strengths of our scoping review are the systematic database searching, the broad inclusion criteria and including an expert consultation stage in a thorough methodological approach to scoping our topic. The work is also timely in view of substantial funding in the UK to support the Farr Institute and the ADRN and collaborative working between the two, which is likely to enhance the potential for developing cross-sectoral linkage studies.[35,36] A limitation of this literature review is that it may have missed routine secondary uses of healthcare data for management, planning, finance and audit purposes, which are taking place within healthcare settings but which would not necessarily be published in the literature.

CONCLUSIONS

Distinguishing between patient-identifiable data and deidentified datasets can help improve conceptual clarity with respect to secondary use policy and planning deliberations in the UK. Innovative secondary uses of data for research purposes hold the promise of new medical knowledge derived from health dataset linkage studies, advances in personalised precision medicine and the advent of cross-sectoral evidence-based policymaking and policy evaluations. In developed nations, domain headings for the various secondary uses of health-related digital data may have attained a level of stability for the foreseeable future and hence only require future updating scoping reviews at longer intervals.

Conflict of interests

The authors have declared no conflicts of interest.

REFERENCES

1. Jones SS, Heaton PS, Rudin RS and Schneider EC. Unravelling the IT productivity paradox – lessons for health care. *New England Journal of Medicine* 2012;366:24. http://dx.doi.org/10.1056/NEJMp1204980. PMid:22693996.

2. Mandl K and Kohane IS. Escaping the EHR trap – the future of health IT. *New England Journal of Medicine* 2012;366:24. http://dx.doi.org/10.1056/NEJMp1203102. PMid:22693995.

3. Black A, Car J, Pagliari C, Anandan C, Cresswell K, Bokun T et al. The Impact of eHealth on the quality and safety of health care: a systematic overview. *PLoS Medicine* 2011;8:e1000387. http://dx.doi.org/10.1371/journal.pmed.1000387. PMid:21267058; PMCid:PMC3022523.

4. Sheikh A, Cornford T, Barber N, Robertson A, Cresswell K, Takian A et al. Implementation and adoption of nationwide electronic health records in secondary care in England: final qualitative results from a prospective national evaluation in 'early adopter' hospitals. *BMJ* 2011;343:d6054. http://dx.doi.org/10.1136/bmj.d6054. PMid:22006942; PMCid:PMC3195310.

5. Cresswell K and Sheikh A. Effective integration of technology into health care needs adequate attention to sociotechnical processes, time and a dose of reality. *Journal of the American Medical Informatics Association* 2012;307:2255–2255.

6. Avery A, Rodgers S, Cantrill JA, Armstrong S, Elliott R, Howard R et al. PINCER trial: a cluster randomized trial comparing the effectiveness and cost-effectiveness of a pharmacist-led IT-based intervention with simple feedback in reducing rates of clinically important errors in medicines management in general practices. *Lancet* 2012;379(9823):p1310–1319; http://dx.doi.org/10.1016/S0140-6736(11)61817-5.

7. Hippisley-Cox J, Coupland C, Vinogradova Y, Robson J, Minhas R, Sheikh A et al. Predicting cardiovascular risk in England and Wales: prospective derivation and validation of QRISK2. *BMJ* 2008;336:1475–1482. http://dx.doi.org/10.1136/bmj.39609.449676.25. PMid:18573856; PMCid:PMC2440904.

8. Lidegaard Ø, Løkkegaard E, Jensen A, Skovlund CW and Keiding N. Thrombotic stroke and myocardial infarction with hormonal contraception. *New England Journal of Medicine* 2012;366:2257–66. http://dx.doi.org/10.1056/NEJMoa1111840. PMid:22693997.

9. Cresswell K and Sheikh A. Electronic health record technology. *Journal of the American Medical Association* 2012 Jun 6;307(21):2255–6. http://dx.doi.org/10.1001/jama.2012.3520. PMid:22706825.

10. American Medical Informatics Association. Secondary uses and re-uses of healthcare data: taxonomy for policy formulation and planning. Available from: http://www.amia.org/sites/amia.org/files/2007-Policy-Meeting-amia-taxonomy-Secondary-Data-Use-Version.pdf. Accessed 15 January 2013.

11. Arksey H and O'Malley L. Scoping studies: towards a methodological framework. *International Journal of Social Research Methodology* 2005;8(1):19–32. http://dx.doi.org/10.1080/1364557032000119616.

12. Levac D, Colquhoun H and O'Brien KK. Scoping studies: advancing the methodology. *Implementation Science* 2010,5:69. http://dx.doi.org/10.1186/1748-5908-5-69.   PMid:20854677; PMCid:PMC2954944.

13. Braun V and Clarke V. Using thematic analysis in psychology. *Qualitative Research in Psychology* 2006;3(2):77–101. http://dx.doi.org/10.1191/1478088706qp063oa.

14. Arendrup M C, Bruun B, Christensen JJ, Fuursted K, Johansen HK, Kjaeldgaard P et al. National surveillance of fungemia in Denmark (2004 to 2009). *Journal of Clinical Microbiology* 2011;49 (1):325–334. http://dx.doi.org/10.1128/JCM.01811-10. PMid:20980569; PMCid:PMC3020479.

15. Colvin L., Slack-Smith L, Stanley FJ and Bower C. Are women with major depression in pregnancy identifiable in population health data? *BMC Pregnancy Childbirth* 2013;13:63. http://dx.doi.org/10.1186/1471-2393-13-63.   PMid:23497210; PMCid:PMC3602106.

16. Schreiweis B, Gruber G and Bergh B. First experiences in implementing a software-based support for patient recruitment at Heidelberg university hospital. *Studies in Health Technology and Informatics* 2012;180:1147–1149. PMid:22874382.

17. Haley, V. B., C. van Antwerpen, et al. Use of administrative data in efficient auditing of hospital- acquired surgical site infections, New York State 2009–2010. *Infection Control and Hospital Epidemiology* 2012;33(6): 565–571. http://dx.doi.org/10.1086/665710. PMid:22561711.

18. Anderson, M, Zito D and Kongable G. Benchmarking glucose results through automation: the 2009 Remote Automated Laboratory System Report. *Journal of Diabetes Science and Technology* 2012;4(6):1507–1513. http://dx.doi.org/10.1177/193229681000400627.

19. Bottacchi E, Corso G, Tosi P, Morosini MV, De Filippis G, Santoni L et al. The cost of first-ever stroke in Valle d'Aosta, Italy: linking clinical registries and administrative data. *BMC Health Services Research* 2012;12:372. http://dx.doi.org/10.1186/1472-6963-12-372. PMid:23110322; PMCid:PMC3507771.

20. Klein D, Staples J, Pittman C and Stepanko C. Using electronic clinical practice audits as needs assessment to produce effective continuing medical education programming. *Medical Teacher* 2012;34(2):151–154. http://dx.doi.org/10.3109/0142159X.2012.644826. PMid:22288993.

21. Roos NP, Black CD, Frohlich N, Decoster C, Cohen MM, Tataryn DJ et al. A population-based health information system. *Medical Care* 1995;33: DS13–20. http://dx.doi.org/10.1097/00005650-199512001-00005. PMid:7500666.

22. Baldwin JA. The Oxford Record Linkage Study as a medical information system. *Journal of Epidemiology and Preventive Medicine* 1972;65:237–9.

23. Kendrick S and Clarke J. The Scottish Record Linkage System. *Health Bull* 1993;51(2):72.

24. Kurland LT and Molgaard CA. The patient record in epidemiology. *Scientific American* 1981;245:54–63. http://dx.doi.org/10.1038/scientificamerican1081-54.

25. British Columbia Centre for Health Services and Policy Research. BC Linked Health Database. Available at: http://www.chspr.ubc.ca/data. Accessed October 2007.

26. Holman CD, Bass AJ, Rouse IL and Hobbs MS. Population-based linkage of health records in Western Australia: development of a health services research linked database. *Australian and New Zealand Journal of Public Health* 1999;23:453-9. http://dx.doi.org/10.1111/j.1467-842X.1999.tb01297.x. PMid:10575763.

27. Mackay D, Haw S, Ayres MD, Fischbacher C and Pell JP. Smoke-free legislation and hospitalizations for childhood asthma. *New England Journal of Medicine* 2010;363:1139–1145. http://dx.doi.org/10.1056/NEJMoa1002861.

28. Risgaard B, Nielsen JB, Jabbari R, Haunsø S, Holst AG, Winkel BG et al. Prior myocardial infarction in the young: predisposes to a high relative risk but low absolute risk of a sudden cardiac death. *Europace* 2013;15(1):48–54. http://dx.doi.org/10.1093/europace/eus190. PMid:22753866.

29. Alan J, Burmas M, Preen D and Pfaff J. Inpatient hospital use in the first year after release from prison: a Western Australian population-based record linkage study. *Australian and New Zealand journal of Public Health* 2011;35(3):264–269. http://dx.doi.org/10.1111/j.1753-6405.2011.00704.x. PMid:21627727.

30. Villanueva K, Pereira G, Knuiman M, Bull F, Wood L, Christian H et al. The impact of the built environment on health across the life course: design of a cross-sectional data linkage study. *BMJ Open* 2013;3(1):e002482. http://dx.doi.org/10.1136/bmjopen-2012-002482.

31. Evans J, Donnelly L, Emslie-Smith A, Alessi D and Morris A. Metformin and reduced risk of cancer in diabetic patients. *BMJ* 2005;330:1304–5. http://dx.doi.org/10.1136/bmj.38415.708634.F7. PMid:15849206; PMCid:PMC558205.

32. Thomas K, Martin R, Davies N, Metcalfe C, Windmeijer F and Gunnell D. Smoking cessation treatment and risk of depression, suicide, and self-harm in the Clinical Practice Research Datalink: prospective cohort study. *BMJ* 2013;347. doi: http://dx.doi.org/10.1136/bmj.f5704.

33. Fone D, Dunstan F, White J, Morgan J, Fry R, Rodgers S et al. Change in alcohol outlet density and alcohol-related harm to public health (CHALICE). *BMC Public Health* 2012;12:428. http://dx.doi.org/10.1186/1471-2458-12-428.   PMid:22691534; PMCid:PMC3409073.

34. Mackay D, Haw S, Ayres J, Fischbacher C and Pell J. Smoke-free Legislation and Hospitalizations for Childhood Asthma. *New England Journal of Medicine* 2010;363:1139–45. http://dx.doi.org/10.1056/NEJMoa1002861. PMid:20843248.

35. The Farr Institute. New job opportunities within Farr London. http://www.farrinstitute.org/1_about.html. Accessed 07 July 2014.

36. The Administrative Data Research Network. http://adrn.ac.uk. Accessed 21 July 2014.

APPENDIX 1

Search strategy 1, MEDLINE format

1. (eHealth or e health or e-health).mp.
2. Telemedicine/
3. Telehealth.mp.
4. Telehealthcare.mp.
5. Telecare.mp.
6. Electronic Prescribing or ePrescribing.mp or e-prescribing.mp.
7. Electronic prescri*.mp.
8. Hospital information systems or information systems or medical records systems, computerised/
9. Health information technology.mp. or Medical Informatics/
10. Medical information system*.mp.
11. Health information system*.mp.
12. Health informatics.mp.
13. Management information systems or management information system*.mp.
14. Integrated advanced information management.mp.
15. Electronic health records
16. Computerised patient record*.mp.
17. Personal health record*.mp. or health records, personal/
18. Decision making, computer-assisted/ or decision support systems, management/ or decision support techniques/ or decision support systems, clinical/
19. Decision support system*.mp.
20. Computerised decision support.mp.
21. Computerised order entry.mp.
22. Electronic patient record.mp.
23. Computerised decision support system*.mp.
24. Medical order entry systems or medical records systems, computerised/
25. Computerised physician order entry.mp.
26. Computerised physician order entry system*.mp.
27. Computerised provider order entry.mp.
28. (Picture archiving and communication system*).mp.
29. or/1-28
30. Public health/
31. Public health informatics
32. Secondary Use*.mp.
33. (quality management and analysis system*).mp.
34. Quality improvement.mp. or quality improvement
35. (Quality control, healthcare or quality control, health care or quality control or healthcare).mp.
36. (Quality indicator*, healthcare or quality indicator*, health care or quality indicator*, health-care).mp.
37. Risk management/
38. Data mining/
39. Data repositor*.mp.
40. Disease regist*.mp.
41. Pharmacovigilance/
42. e-Science.mp.
43. Service design.mp.
44. Service re-design.mp.
45. Commission*.mp.
46. Disease surveillance.mp.
47. Health services research/
48. Health service monitoring.mp.
49. Quality Assurance, Health Care/
50. Health surveillance.mp.
51. Epidemiology/
52. Clinical Coding/
53. (Clinical audit or medical audit).mp.
54. Data linkage.mp.
55. Finance*.mp.
56. or/30-55
57. 29 and 56
58. Limit 57 to yr="2000-current"
59. advertisements/ or animation/ or architectural drawings/ or bibliography/ or biography/ or book illustrations/ or bookplates/ or charts/ or comment/ or letter/ or news/ or patient education handout/ or published erratum/ or "retraction of publication"/
60. 58 not 59

APPENDIX 2

Search strategy 2, free-field format

(eHealth or e health or e-health or telemedicine or telehealth or telecare or telehealthcare or electronic prescri* or e-prescribing or eprescribing or health information technology or medical informatics or medical information system* or health information system* or health informatics or computerised medical records system* or hospital information system* or management information system* or electronic health record* or computerised patient record* or personal health record* or decision support system* or clinical decision support system* or computerised decision support or computerised order entry or electronic patient record* or medical order entry system* or computerised physician order entry system* or computerised provider order entry or picture archiving and communication system*)

AND:

(Public health or public health informatics or secondary use* or quality management and analysis system* or quality improvement or quality of health care or quality of healthcare or health care quality control or healthcare quality control or health care quality indicator* or healthcare quality indicator* or risk management or data mining or data repository or disease regist* or pharmacovigilance or e-science or service design or commission* or disease surveillance or health service* research or health service* monitoring or health care quality assurance or healthcare quality assurance or health surveillance or epidemiology or clinical coding or clinical audit or medical audit or data linkage or finances)

APPENDIX 3

Long list of secondary uses identified from scoping review

Audit (e.g. diabetes care per guidelines; statin use; screening uptake)

Bio surveillance for disease outbreaks

Cause of death status for policy/planning in low to middle income countries (e.g. Ramatige 2012)

Clinical tool development (e.g. mammography images)

Clinical workflow templates

Communication between specialists, payers and providers, etc.

Court reports on defendants

Crisis/risk/disaster management; strategic decision making for data mining–commercial text mining tools for data mining–compliance patterns; Decision Support development (e.g. toxicology)

Decreasing costs by reducing medical errors, duplication of services Development of integrated care pathways Disease profiling

Disease registries (e.g. stroke, lupus)

Disease surveillance

Education (teaching medical students, healthcare professionals, pharmacists, etc.)

Epidemiology

Finance–billing

Finance–cost benefit analyses

Finance–performance based reimbursement

Geographical mapping–for healthcare and policy

Health literacy and empowerment; evidence developmen

Healthcare safety and guality

Return on investment

HITs–a platfbrm/fbundation for strategic guidelines/next generation comparative effectiveness research in real world settings Hospital readmission risk calculation (e.g. heart failure)

Human phenome database (dermatology) (leaming/education and research)

Identifying and targeting at-risk patient populations Immigrants/healthcare Improvement of data standards Incident analysis (radiotherapy)

Indicator development

Infectious disease/disaster management

Integration of clinical care and administrative data; linkage of data; data mining management/monitoring outcome tracking

Patient self-management and self-efficacy; personalised medicine performance indicators (e.g. surgery)

Performance monitoring

Pharmacovigilance/pharmacoepidemiology (adverse events)

Predictive models for disease outbreaks (e.g. measles)

Preventative care and health promotion Prognostic marker

Prognostic modelling e.g. coronary artery disease (CALIBER study, Rapsomaniki, Shah et al., 2012)

Public health records

Public health surveillance

Quality control of services (e.g. ECT)

Quality improvement (Comparative effectiveness research, clinical decision support and disease surveillance)

Quality measurement

Race/ethnicity/sex differences for risk (e.g. REGARD)

Reducing conventional administrative work for intractable/notifiable diseases Reporting and monitoring Research–Case control studies Research–Cohort studies

Research–disease markers/neurodegenerative disorders (neuGRID, Redolfi et al., 2009, research–distributed virtual global lab (Suominen, 2012)

Research–Electronic Health Records for Clinical Research collaboration; see Thomson et al., 2012

Research–Genetic research/disease traits

Research–identifying potential research participants from particular populations Research–observational epidemiological studies (e.g. Rochester Epidemiology Project)

Research–pathology

Research–prescription abandonment/treatment adherence (e.g. Shrank, 2010)

Research–Rare disorders research (ORPHANET)

Research/Natural Language Processing (NLP), e.g. to assess smoking Resource allocation

Scales validation (e.g. asthma control scale)

Service planning

Standards development (e.g. for EHRs)

Toxicovigilance–Rapid Alert System for Chemicals database tracking of trends and prediction of patterns