Measuring Quality of Healthcare Outcomes in Type 2 Diabetes from Routine Data: a Seven-nation Survey Conducted by the IMIA Primary Health Care Working Group

Primary Health Care Informatics Working Group Contribution to the Year Book of Medical Informatics 2017

W. Hinton[1], H. Liyanage[1], A. McGovern[1], S.-T. Liaw[2], C. Kuziemsky[3], N. Munro[1], S. de Lusignan[1]

[1] Clinical Informatics & Health Outcomes Research Group, Department of Clinical & Experimental Medicine, University of Surrey, Guildford, Surrey, UK

[2] School of Public Health & Community Medicine, UNSW Medicine, Australia

[3] Telfer School of Management, University of Ottawa, Ottawa, Ontario, Canada

Summary

Background: The Institute of Medicine framework defines six dimensions of quality for healthcare systems: (1) safety, (2) effectiveness, (3) patient centeredness, (4) timeliness of care, (5) efficiency, and (6) equity. Large health datasets provide an opportunity to assess quality in these areas.

Objective: To perform an international comparison of the measurability of the delivery of these aims, in people with type 2 diabetes mellitus (T2DM) from large datasets.

Method: We conducted a survey to assess healthcare outcomes data quality of existing databases and disseminated this through professional networks. We examined the data sources used to collect the data, frequency of data uploads, and data types used for identifying people with T2DM. We compared data completeness across the six areas of healthcare quality, using selected measures pertinent to T2DM management.

Results: We received 14 responses from seven countries (Australia, Canada, Italy, the Netherlands, Norway, Portugal, Turkey and the UK). Most databases reported frequent data uploads and would be capable of near real time analysis of healthcare quality. The majority of recorded data related to safety (particularly medication adverse events) and treatment efficacy (glycaemic control and microvascular disease). Data potentially measuring equity was less well recorded. Recording levels were lowest for patient-centred care, timeliness of care, and system efficiency, with the majority of databases containing no data in these areas. Databases using primary care sources had higher data quality across all areas measured.

Conclusion: Data quality could be improved particularly in the areas of patient-centred care, timeliness, and efficiency. Primary care derived datasets may be most suited to healthcare quality assessment.

Keywords

Data quality; data accuracy; diabetes mellitus; type 2; medical record systems

Yearb Med Inform 2017:201-8
http://dx.doi.org/10.15265/IY-2017-005
Published online May 8, 2017

Introduction

Computerised medical records (CMRs) whilst originally introduced to improve quality [1,2], now provide large volumes of data that can be used to assess and compare the quality of healthcare on a global scale. Although several frameworks for healthcare quality assessment have been developed, one of the most influential frameworks is that set out by the US Institute of Medicine (IOM) [3]. This framework defines six aims for healthcare systems (Box 1). CMRs potentially provide the ability to rapidly amalgamate and analyse data on a large scale to assess quality [4].

An aging population has an increased prevalence of chronic illness, of which type 2 diabetes mellitus (T2DM) is an important condition. The global prevalence of T2DM is continuing to rise, and the condition is now considered a global pandemic [5]. Developing and disseminating evidence to support chronic disease management (CDM) is a significant challenge for several reasons including the need to collect and integrate multiple disparate data sources while also needing timely turnaround of findings to support patient care delivery [6]. Whilst high quality data is needed to support ongoing analyses as part of improving health outcomes, in reality, the collection, analysis, and re-circulation of findings is often done in a piecemeal fashion and not as part of a well-defined framework.

The IMIA Primary Health Care Informatics Working Group (PHCIWG) has a long interest in data quality. Its work has included defining key concepts to assess readiness of data for use in research [7] and systematic methods to underpin pooled use of data and interoperability [8]. A comprehensive realist review of the data quality literature helped to determine prerequisites for using routine data for integrated care [9] and to support practice-based research networks [10]. Strategies included structured data quality reports and feedback sessions [11] and working with unstructured CMR

data [12, 13]. Research was also conducted to assess the effect of data quality on the development of clinical predictive models and the identification and development of cohorts and registries, using both structured [14] and unstructured [15] data.

This paper presents the results of an investigation we carried out to explore how the comparison of routinely collected data could allow the comparison of the quality of care in T2DM across various countries.

Methods

We provide a two-component analysis of global data quality. Firstly, we undertook a literature review of data quality research with a focus on data quality in diabetes. Secondly, we undertook a survey of working group members to give a contemporary cross-section of data quality in existing databases about the quality of care of people with T2DM.

Literature Review

We carried out a literature review to identify published research related to data quality in studies related to T2DM patients. We searched PUBMED/Medline, Scopus, Web of Science, CINAHL, and the Cochrane Database for publications related to this topic. The search terms used comprised "data quality" (or "data accuracy") and "diabetes".

An overview of the literature review is given in the adapted PRISMA flow chart shown in Figure 1. Research papers published between November 2006 and November 2016 were included to provide an overview of the last ten years of literature. We limited the literature search to publications written in the English language. We included relevant studies from the references cited within the studies returned by the search and additional studies known to us from our experience in this domain.

The initial search yielded a total of 321 publications, which was then reduced to 263 after removing duplicates. By conducting a title and abstract review, we further reduced the number of relevant studies to 25. From the analysis of references and the addition of studies we were previously aware of, three studies were added. These 28 publications were used for the final in-depth review.

We provide a narrative summary on what is known about diabetes data quality in existing databases on the basis of the papers identified by this search strategy. We also provide comments on the limitations of the current literature and suggestions for future research in this area.

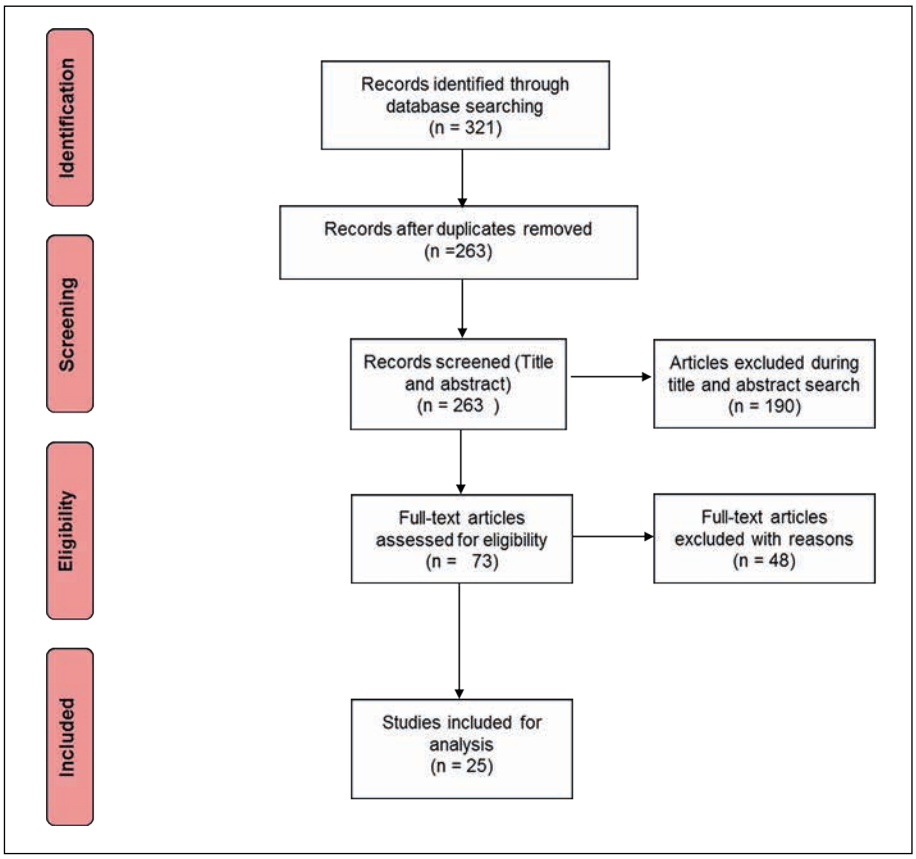

Fig. 1 Adapted PRISMA flow diagram of the literature search

Box 1 The six areas of healthcare quality as suggested by the IOM [3]

1. Safe: avoidance of patient harm
2. Effective: evidence based-practice
3. Patient centred: care tailored towards the individual
4. Timely: reduction of harmful delays
5. Efficient: avoidance of waste
6. Equitable: providing the same care quality to all

Survey for Assessing Diabetes Data Quality and Availability

We conducted a survey analysis of T2DM data quality, collected and recorded in existing large datasets from various countries. We surveyed data recording across the six areas of healthcare quality defined by the IOM (Box 1).

Using a preformatted data collection form (Appendix 1), we collected data for each database on database country, information source, upload frequency, funding source, and denominator population. We also queried the method used by database analysts to identify T2DM cases. For each areas of healthcare quality we carefully selected one or two key outcome measures relevant to the care of T2DM based on national and international diabetes

guideline recommendations. We chose retinopathy as a marker of intervention efficacy (rather than other diabetes outcomes) as this is the key microvascular diabetes outcome which has historically determined diabetes targets [16, 17]. We distributed the data collection form by email to healthcare database representatives and diabetes academics identified via various professional networks. In the case the recipients reached were not the right persons to answer the survey, we asked that they forward the email onto someone better placed to complete the survey, known as snowball sampling [18]. We distributed 35 surveys in total.

We provide a summary of the survey results, including the proportion of incomplete data in responses.

Results

Literature Review

We found limited evidence in specific studies focusing on quality of data for diabetes research. Most work was in relation to generic data quality metrics and assessment where diabetes was considered as a comorbidity or a risk factor. We found two overarching themes during the detailed analysis. The first theme was related to the challenges and issues in collecting data (CMR extraction, data quality), while the second theme was about using the data to monitor disease management.

Goudswaard et al. conducted a cross-sectional survey in general practitioner (GP) practices in the Netherlands and were unable to evidence an association between incomplete clinical records and glycaemic control (although they reported that low levels of recording affected the systematic delivery of care) [19]. Keating et al. assessed quality indicators using administrative and medical records from a cohort of diabetic patients [20]. They reported that tests such as HbA1c and LDL cholesterol have a higher recording in CMR data while retinopathy screening was more often recorded in administrative data. Richesson et al. conducted a study to compare phenotype definitions for diabetes considering seven different phenotype definitions [21].

Two studies analysed data quality within a UK-based research and surveillance primary care network [22, 23]. They reported a high level of data recording in people with diabetes, such as ethnicity identification (82.1%), smoking status (99.3%), alcohol use (93.3%), glycated haemoglobin (HbA1c; 97.9%), body mass index (98.0%), blood pressure (99.4%), cholesterol (87.4%), and renal function (97.8%).

Epidemiology cohorts focused on diabetes have provided better quality datasets for research by assembling data from multiple data sources (such as clinical records, prescriptions, laboratory tests, and mortality files). Linkage across multiple dimensions of care has successfully increased the quality of care [24]. Furthermore, validation studies have demonstrated that data related to processes and outcomes of diabetic care is well recorded within primary care [25]. Bailie et al. have demonstrated that the consistency of denominator data is important when comparing indicators over time and between services [26].

Governance of data networks and health information sharing were important and constrained data use.

Survey Results

We received 14 responses from seven countries (Australia, Canada, Italy, the Netherlands, Norway, Portugal, Turkey, and the United Kingdom), a response rate of 40%. The most common data source among the databases surveyed was primary care health records (Figure 2). All databases surveyed were actively updated, most of them on a daily basis (Figure 3).

A wide variety of data types has been used to identify people with T2DM from these databases (Figure 4). The majority of databases used multiple data types (Table 1) to identify diabetes (three databases used two data types, two databases used three data types, and five databases used four data types). The population denominator of responding databases included four national databases, four practice networks, four locality-specific databases, and two disease registers.

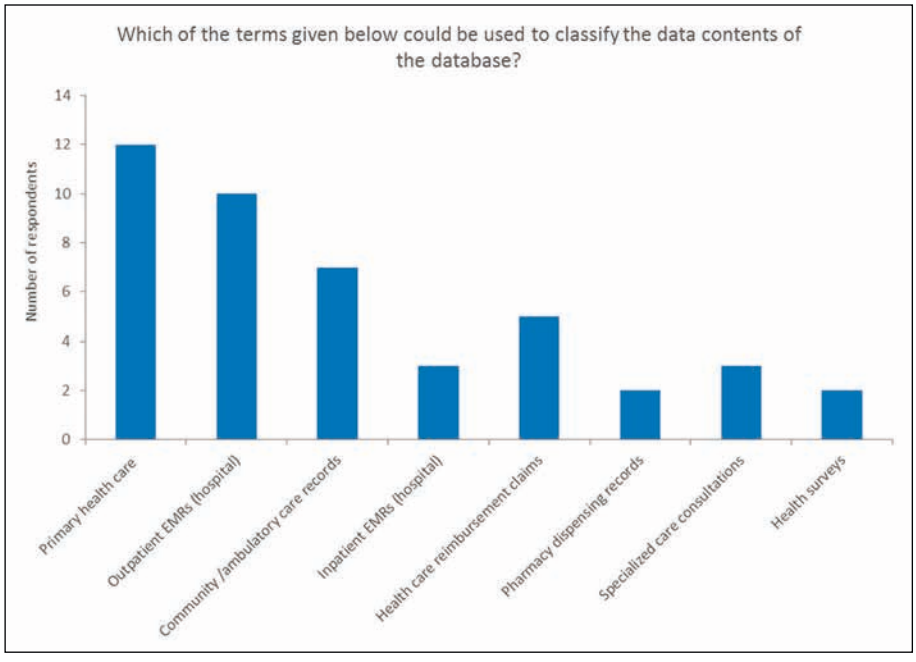

Fig. 2 Graph – Contents of databases (sent by responders) (Survey Question 4)

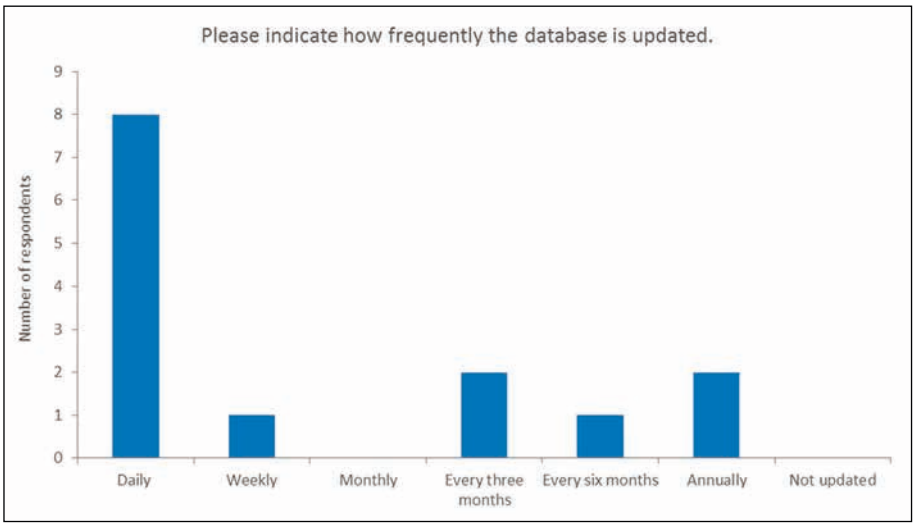

Fig. 3 Graph – Frequency of update (Survey Question 5)

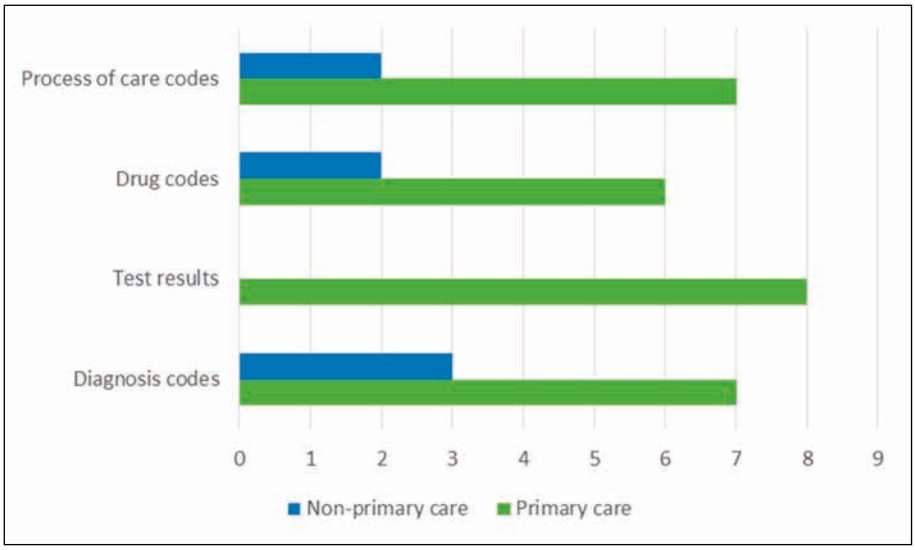

Fig. 4 Methods used for identifying T2DM (Survey Question 10)

The recording of data within the selected healthcare quality areas was highly variable across the databases surveyed. In general, data recording was better in data from primary care sources than in data from other sources (Figure 5). Recording within the areas related to patient-centred care, timeliness of care, and system efficiency was lowest. Of the 14 databases surveyed, only one had partial or complete data recording in every healthcare quality area assessed.

Discussion

Principal Findings

We have identified a number of large databases which could be used to compare healthcare quality in T2DM across various countries. Most of these databases reported very frequent data uploads and would be capable of real time or near-real time analysis of healthcare quality. The majority of databases recorded data related to safety (particularly medication adverse events) and treatment efficacy (glycaemic control, microvascular disease, and retinopathy). This may potentially be owing to the data being recorded numerically within the CMRs. Data which could be used to measure equity (patient ethnicity and socioeconomic status) were less recorded (just in over 50% of databases). The lowest levels of recording were observed in the areas of patient-centred care, timeliness of care, and system efficiency with the majority of databases containing no data in these areas. These areas are multidimensional as opposed to data associated to hypoglycaemia or glaucoma. Databases using primary care CMR data had higher data quality across the areas measured than those from other sources. Timeliness of care may be simpler to derive from recorded data than to be recorded as such into CMRs. Through these findings, we propose a taxonomy considering the dimensionality of the measures: single and simple measures, derived measures (such as timeliness of care), multidimensional measures (such as an index of patient-centred care or a depression scale or a Bartel's index) [27].

Implications of Findings

Whilst healthcare databases provide an excellent opportunity to monitor healthcare quality due to their large size and the high frequency data collection, they were not designed with this purpose in mind. Data recording presently appears to fit the traditional medical model [28] with data recording focused on the measurement of disease-related parameters rather than on other important healthcare quality outcomes: clear documentation of patient-centred targets and societal considerations is limited in existing datasets. Almost universally, most recent T2DM guidelines emphasize the importance of individualised care [29-31]. So this is perhaps a surprising, if not disappointing, finding.

Hypoglycaemia was not at all recorded in a large proportion of databases. This is an extremely important clinical safety issue in people with diabetes which merits close clinical scrutiny [32]. The absence of data recording in this area is a concern.

Table 1 Methods used for identifying T2DM by respondent (Survey Question 10)

Respondent	Source	Process of care codes	Drug codes	Test results	Diagnosis codes	Number of data types
1	Non-primary care	0	1	0	1	2
2	Non-primary care	0	0	0	0	0
3	Non-primary care	1	0	0	1	2
4	Non-primary care	0	1	0	0	1
5	Non-primary care	1	0	0	0	1
6	Non-primary care	0	0	0	1	1
7	Primary care	1	1	1	1	4
8	Primary care	1	0	1	1	3
9	Primary care	1	1	1	1	4
10	Primary care	0	0	1	1	2
11	Primary care	1	1	1	1	4
12	Primary care	1	1	1	1	4
13	Primary care	1	1	1	0	3
14	Primary care	1	1	1	1	4

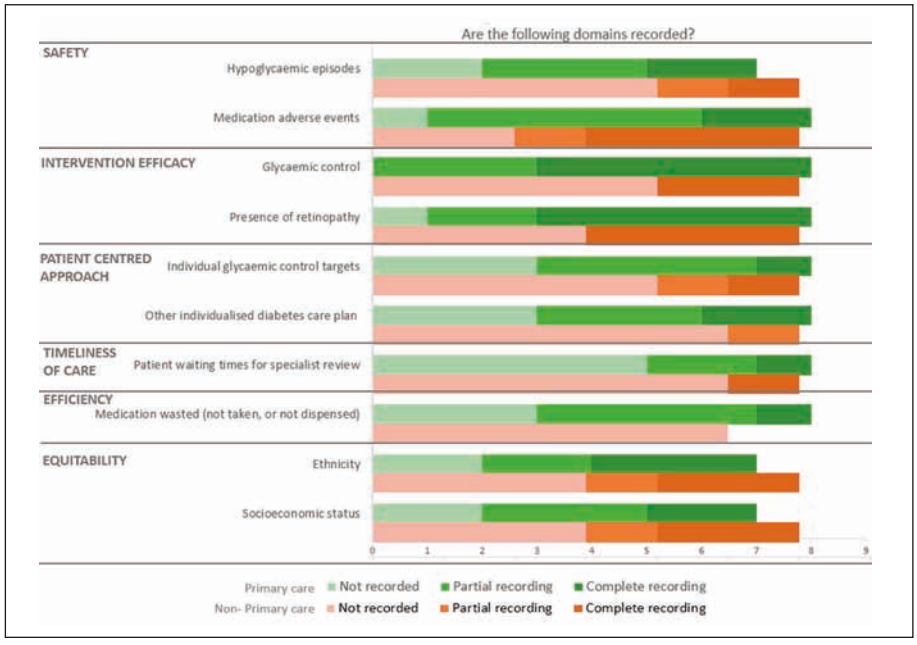

Fig. 5 Areas being recorded and the level of recording (Survey Question 11)

These results seem to indicate that primary care sourced data has better levels of recording across all health care quality areas when compared to other data sources, and it is likely that primary care data provides better opportunities for monitoring healthcare outcomes globally. However, further studies need to be conducted in order to validate these findings.

In addition to the findings from the survey, CMRs can be designed to calculate individual risks such as multi-morbidity, polypharmacy, and other elements important in individual risk assessment (e.g. absolute cardiovascular risk calculation).

Comparison with the Literature

In our literature review, we found few specific studies focusing on quality of data for diabetes research. Most works were in relation to generic data quality metrics and assessments where diabetes was considered as a comorbidity or a risk factor. Large scale studies of diabetes related data quality are needed. To the best of our knowledge this is the first assessment of the quality of diabetes data from large datasets provided by different countries. Beyond this initial analysis, further work is needed to quantify data quality in each key area directly from these databases.

Limitations of the Method

The limitations of this analysis are primarily those of survey studies: possible bias caused by incomplete responses to questionnaires and the limitations of self-reported data quality (by those with a role as database representatives). We were also probably only aware of a small selection of all the healthcare databases which exist globally, and therefore our survey provides data only from a small selection of world-wide databases. In addition, we did not conduct a systematic assessment of whether the databases used are comparable with respect to data quality and population demographics. Nevertheless, given these limitations these data provide a starting point for further analyses in this area and a framework for assessing data quality of databases for their value in assessing healthcare quality.

Conclusions

Whilst large databases have great potential for monitoring healthcare quality, the recording of healthcare outcomes data in a number of key areas requires considerable improvement internationally. The areas requiring improved data collection include patient-centred care, timeliness of care, and system efficiency. Improved data quality in these areas would facilitate local healthcare quality improvement

and comparison of quality nationally and internationally. Despite these limitations, primary care derived datasets showed high levels of data recording across all domains of healthcare outcomes and therefore may provide the best data sources for healthcare quality analysis.

Acknowledgements

This paper is an academic exercise by the Primary Health Care Informatics Working Group of IMIA. We would like to acknowledge Oliver Frank, Pinar Topsever, Ana Gabriela Ribeiro, Ana Maria, Tania Stafinski, Heidi Lyshol, Robert Verheij, Lara Tramontan, Selam Mequanint, Stewart Harris, and Catherine Pelletier for their assistance in completing the survey. This paper and its contributions are the work of volunteers.

Funding

This paper was written by volunteer authors from the working group.

References

1. Majeed A, Car J, Sheikh A. Accuracy and completeness of electronic patient records in primary care. Fam Pract 2008;25(4):213-4.
2. Clay RA. The advantages of electronic health records. Electronic records can improve patient care. Here's how. Available from: http://www.apa.org/monitor/2012/05/electronic-records.aspx. [Accessed 12th December 2016].
3. Institute of Medicine (IOM). Crossing the Quality Chasm: A New Health System for the 21st Century. Washington, D.C: National Academy Press; 2001.
4. de Lusignan S, van Weel C. The use of routinely collected computer data for research in primary care: opportunities and challenges. Fam Pract 2006 Apr;23(2):253-63.
5. The diabetes pandemic. Lancet 2011 Jul 9;378(9786):99.
6. Liyanage H, Liaw ST, Kuziemsky C, Terry AL, Jones S, Soler JK, et al. The Evidence-base for Using Ontologies and Semantic Integration Methodologies to Support Integrated Chronic Disease Management in Primary and Ambulatory Care: Realist Review. Contribution of the IMIA Primary Health Care Informatics WG. Yearb Med Inform 2013;:147-54.
7. de Lusignan S, Liaw ST, Krause P, Curcin V, Vicente M, Michalakidis G, 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. Yearb Med Inform 2011;:112-21.
8. de Lusignan S, Liaw S, Michalakidis G, Jones S. Defining data sets and creating data dictionaries for quality improvement and research in chronic disease using routinely collected data: an ontology driven approach. Inform Prim Care 2011;19(3):127-34.
9. Liaw S, Rahimi A, Ray P, Taggart J, Dennis S, de Lusignan S, et al. Towards an ontology for data quality in integrated chronic disease: a realist review of the literature. Int J Med Inform 2013;82(1):10–24.
10. Taggart J, Liaw ST, Dennis S, Yu H, Rahimi A, Jalaludin B, et al.The University of NSW electronic practice based research network: disease registers, data quality and utility. Stud Health Technol Inform 2012; 178:219-27.
11. Taggart J, Liaw S-T, Yu H. Structured data quality reports to improve EHR data quality. Int J Med Inform 2015;84:1094-8.
12. Jonnagaddala J, Liaw S, Ray P, Kumar M, Chang N, Dai H. Coronary artery disease risk assessment from unstructured electronic health records using text mining. J Biomed Inform 2015;58 Suppl:S203-10.
13. Jonnagaddala J, Liaw S, Ray P, Kumar M, Dai H, Hsu C. Identification and Progression of Heart Disease Risk Factors in Diabetic Patients from Longitudinal Electronic Health Records. Biomed Res Int 2015;2015:636371.
14. Liaw S, Taggart J, Yu H, de Lusignan S, Kuziemsky C, Hayen A. Integrating electronic health record information to support integrated care: practical application of ontologies to improve the accuracy of diabetes disease registers. J Biomed Inform 2014;52:364-72.
15. Rahimi A, Liaw S, Ray P, Taggart J, Yu H. Ontological specification of quality of chronic disease data in EHRs to support decision analytics: a realist review. Decision Analytics 2014;1(1):5.
16. Stratton, I.M., et al., Association of glycaemia with macrovascular and microvascular complications of type 2 diabetes (UKPDS 35): prospective observational study. BMJ 2000 Aug 12;321(7258):405-12.
7. Kohner EM. Microvascular disease: what does the UKPDS tell us about diabetic retinopathy? Diabet Med 2008;25 Suppl 2:20-4.
18. Goodman Leo A. Snowball Sampling. Annals of Mathematical Statistics 1961;32:148–70.
19. Goudswaard AN, Lam K, Stolk RP, Rutten GE. Quality of recording of data from patients with type 2 diabetes is not a valid indicator of quality of care. A cross-sectional study. Fam Pract 2003 Apr;20(2):173-7.
20. Keating NL, Landrum MB, Landon BE, Ayanian JZ, Borbas C, Guadagnoli E. Measuring the quality of diabetes care using administrative data: is there bias? Health Serv Res 2003 Dec;38(6 Pt 1):1529-45.
21. Richesson RL, Rusincovitch SA, Wixted D, Batch BC, Feinglos MN, Miranda ML, et al. A comparison of phenotype definitions for diabetes mellitus. J Am Med Inform Assoc 2013;20:e319–e326.
22. McGovern A, Hinton W, Correa A, Munro N, Whyte M, de Lusignan S. Real-world evidence studies into treatment adherence, thresholds for intervention and disparities in treatment in people with type 2 diabetes in the UK. BMJ Open 2016 Nov 24;6(11):e012801.
23. Correa A, Hinton W, McGovern A, van Vlymen J, Yonova I, Jones S, de Lusignan S, et al. Royal College of General Practitioners Research and Surveillance Centre (RCGP RSC) sentinel network: a cohort profile. BMJ Open 2016 Apr 20;6(4):e011092.
24. Miller DR, Pogach L. Longitudinal approaches to evaluate health care quality and outcomes: the Veterans Health Administration diabetes epidemiology cohorts. J Diabetes Sci Technol 2008 Jan;2(1):24-32.
25. Hill F, Bradley C.A computer based, automated analysis of process and outcomes of diabetic care in 23 GP practices. Ir Med J 2012 Feb; 105(2):45-7.
26. Bailie R, Bailie J, Chakraborty A, Swift K. Consistency of denominator data in electronic health records in Australian primary healthcare services: enhancing data quality. Aust J Prim Health 2015;21(4):450-9.
27. Turnbull PJ, Sinclair AJ. Evaluation of nutritional status and its relationship with functional status in older citizens with diabetes mellitus using the mini nutritional assessment (MNA) tool--a preliminary investigation. J Nutr Health Aging 2002 May;6(3):185-9.
28. Engel G. The need for a new medical model: a challenge for biomedicine. Science 1977;196(4286):129-36.
29. Inzucchi SE, Bergenstal RM, Buse JB, Diamant M, Ferrannini E, Nauck M, et al. Management of Hyperglycemia in Type 2 Diabetes: A Patient-Centered Approach. Position Statement of the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD). Diabetes Care 2012;35(6):1364-79.
30. American Diabetes Association. ADA: Standards of Medical Care in Diabetes 2016. Diabetes Care 2016;39(S1):S1-S112.
31. NICE. Type 2 diabetes in adults: management. NICE guidelines [NG28]2015. Available from: https://www.nice.org.uk/guidance/ng28. [Accessed 12th December 2016].
32. Shafiee G, Mohajeri-Tehrani M, Pajouhi M, Larijani B. The importance of hypoglycemia in diabetic patients. Journal of Diabetes & Metabolic Disorders 2012;11:17.

Correspondence to:
Simon de Lusignan
Professor of Primary Care and Clinical Informatics
Clinical Informatics & Health Outcomes Research Group
Department of Clinical & Experimental Medicine
University of Surrey
Guildford, Surrey GU2 7XH
United Kingdom
E-mail: s.lusignan@surrey.ac.uk

Appendix

IMIA Primary Health Care Informatics Working Group: Survey of Health Data Sources to Support Diabetes Studies

With the global prevalence of type 2 diabetes continuing to rise, accurately measuring the global disease burden from this condition is vital. Large datasets are the ideal way to collate this information but in order to assess their compatibility and comparability an assessment of data quality is needed across these datasets.

Using this survey, we intend to conduct an international comparison of data sources available for diabetes studies. The results of the study will be published in the Yearbook of Medical Informatics, 2017 edition.

DATABASE DESCRIPTION

1. Country: .....
2. Name of the database/register: .....
3. Database/register website (URL): .....

4. Which of the terms given below could be used to classify the data contents of the database? (select all that apply)
   - ☐ Primary health care
   - ☐ Outpatient electronic medical records
   - ☐ Community /ambulatory care records
   - ☐ Inpatient electronic medical records / hospital
   - ☐ Health care reimbursement claims, including date and place of service, patient, diagnoses, treatment.
   - ☐ Pharmacy dispensing records
   - ☐ Specialized care consultations
   - ☐ Specific registry (inc. chronic or rare disease, cancer registries)
   - ☐ Health surveys
   - ☐ Other (please specify) .....

5. Please indicate how frequently the database is updated:
   - ☐ Daily (ongoing data entry)
   - ☐ Weekly
   - ☐ Monthly
   - ☐ Three monthly
   - ☐ Six monthly
   - ☐ Annually
   - ☐ Not updated

6. Funding source for database: .....

POPULATION DENOMINATOR

7. Is the database population from (delete as applicable):
   a. a defined population (Y/N)
   b. a registry based system (Y/N)

8. Population covered: e.g. whole country / defined locality (please specify) ..... / Other (please specify) .....

9. Age range covered: .....

CASE DEFINITION

10. Methods used to identify people with type 2 diabetes (delete as applicable):
    a. Diagnosis codes Y/N
    b. Test results Y/N
    c. Drug codes Y/N
    d. Process of care codes (e.g. attends diabetes clinic) Y/N
    e. Other method Y/N (please specify)
    ……

MEASURES OF HEALTHCARE QUALITY

11. Are the following areas recorded?

	Not recorded	Partial recording	Complete recording
Safety			
1. Hypoglycaemic episodes	☐	☐	☐
2. Medication adverse events	☐	☐	☐
Intervention efficacy			
1. Glycaemic control	☐	☐	☐
2. Presence of retinopathy	☐	☐	☐
Patient centred approach			
1. Individual glycaemic control targets	☐	☐	☐
2. Other individualised diabetes care plan	☐	☐	☐
Timeliness of care			
1. Patient waiting times for specialist review	☐	☐	☐
Efficiency			
1. Medication wasted (not taken, or not dispensed)	☐	☐	☐
Equity			
1. Ethnicity	☐	☐	☐
2. Socioeconomic status	☐	☐	☐

ADDITIONAL INFORMATION

12. Please use the space below for any additional information
    ……
    ……
    ……