Received Date: 17-Dec-2010
Revised Date : 01-Jun-2011
Accepted Date: 09-Sep-2011
Article type : Short Report

Running Head: Automating identification of classification and coding errors in diabetes

Article Type: Short Report Health Economics

Automated identification of miscoded and misclassified cases of diabetes from computer records

Ahmed-Ramadan Sadek¹, Jeremy van Vlymen², Kamlesh Khunti³ and Simon de Lusignan²

¹Division of Population Health Sciences and Education, St. George's – University of London, London, ²Department of Health Care Management and Policy, Faculty of Management and Law, University of Surrey, Guildford, ³Department of Health Sciences, University of Leicester, Leicester, UK

Correspondence to: Simon de Lusignan, Department of Health Care Management and Policy, Faculty of Management and Law, University of Surrey, Guildford GU2 7XH, UK
E-mail: s.lusignan@surrey.ac.uk

Aims To develop a computer processable algorithm, capable of running automated searches of routine data that flag miscoded and misclassified cases of diabetes for subsequent clinical review.

Method Anonymized computer data from the Quality Improvement in Chronic Kidney Disease (QICKD) trial ($n = 942\,031$) were analysed using a binary method to assess the accuracy of data on diabetes diagnosis. Diagnostic codes were processed and stratified into: definite, probable and possible diagnosis of Type 1 or Type 2 diabetes. Diagnostic accuracy was improved by using prescription compatibility and temporally sequenced anthropomorphic and biochemical data. Bayesian false detection rate analysis was used to compare findings with those of an entirely independent and more complex manual sort of the first round QICKD study data ($n = 760\,588$).

Results The prevalence of definite diagnosis of Type 1 diabetes and Type 2 diabetes were 0.32% and 3.27% respectively when using the binary search method. Up to 35% of Type 1 diabetes and 0.1% of Type 2 diabetes were miscoded or misclassified on the basis of age/BMI and coding. False detection rate analysis demonstrated a close correlation between the new method and the published hand-crafted sort. Both methods had the highest false detection rate values when coding, therapeutic, anthropomorphic and biochemical filters were used (up to 90% for the new and 75% for the hand-crafted search method).

Conclusions A simple computerized algorithm achieves very similar results to more complex search strategies to identify miscoded and misclassified cases of both Type 1 diabetes and Type 2 diabetes. It has the potential to be used as an automated audit instrument to improve quality of diabetes diagnosis.

Keywords diabetes mellitus, diagnostic errors, computerized medical informatics, medical records systems.

This is an Accepted Article that has been peer-reviewed and approved for publication in the *Diabetic Medicine*, but has yet to undergo copy-editing and proof correction. Please cite this article as an "Accepted Article"; doi: 10.1111/j.1464-5491.2011.03457.x

Introduction

Primary care records are largely computerised with most primary care clinicians entering data at the point of care. Electronic patient records facilitate improved management of chronic diseases including diabetes [1]. However, a combination of factors may reduce data quality, including: time pressures that limit what is recorded in the electronic patient record; changes in diagnostic criteria and classification rules for diabetes [2]; sampling labelling problems (it may not be obvious if a sample is fasted) [3]; and idiosyncrasy in the coding interface. Furthermore, the various brands of EPR offer different coding choices for the same search term [4].

We have demonstrated that a pragmatic search strategy can identify classification problems in diabetes. Many people labelled as having Type 1 diabetes (Type 1 diabetes) are misclassified and actually have Type 2 diabetes mellitus (Type 2 diabetes) and some with Type 2 diabetes may not have diabetes at all [5]. Coding of diabetes diagnostic data is both complex [6] and in need of refinement [5,7] and current practice probably overestimates the prevalence of diabetes [5]. Recognition of these gaps has led NHS Diabetes [8] and the Royal College of General Practitioners (RCGP) to launch 'The Classification of Diabetes' initiative [9].

The pragmatic method of identifying classification errors cannot easily be converted into code that can be processed by existing electronic patient record systems to automatically identify errors and flag patients for review. The pragmatic method can be used as an audit method, but is labour intensive and clinician time is at a premium. We therefore set out to develop a binary method that would allow automatic identification of misclassified cases for subsequent clinical review.

Methods

We used two datasets from the Quality Improvement in Chronic Kidney Disease (QICKD) trial [10]. The QICKD study is being conducted in 129 practices drawn from localities across England. It is intended to have a population profile to match the National average, although it has a slight excess of working-age people and slightly fewer older people aged 60–75 years. The standardized prevalence of diabetes within the study cohort is 4.0% [11]. The first dataset was collected at baseline and the second set of data was collected approximately 1 year later. There was no change in diabetes prevalence between these collection points. Although these data were extracted for a cluster randomized trial they are from ordinary general practices taking part in a quality improvement trial in chronic kidney disease, and should reflect standard routinely collected primary care data.

We designed and tested a simple structured search engine employing Boolean logic, which might be used to improve data quality and highlight cases requiring further investigation [5]. We used a simple binary algorithm to sort diagnostic codes into definite, probable or possible Type 1 diabetes or Type 2 diabetes based on the specificity of the diagnostic term. In addition we identified whether there were and contradictory codes, (e.g. classified as both Type 1 diabetes and Type 2 diabetes; see the Supporting Information). We then applied filters based on the NHS Diabetes/RCGP algorithm [9,12]. We explored compatibility of insulin and/or oral anti-diabetic drugs were compatible with the particular type of diabetes. We next explored whether age and BMI suggested a classification error.

Using established methods, variables were exported into a flat file, one line per anonymized patient record and one column per variable [13,14]. Our initial sorting strategy grouped people into definite Type 1 diabetes or Type 2 diabetes where they had at least two specific diagnostic codes (in the UK Read codes C10E for Type 1 diabetes and C10F for Type 2 diabetes) over a 4-year period in the absence of any contradictory codes.

Where people had specific codes for Type 1 diabetes or Type 2 diabetes we used binary code to represent the presence or absence of a definite diagnosis in each of the last four years. We then compared the last four years' data for Type 2 diabetes with the same four years' Type 1 diabetes data. To qualify as a definite diagnosis it was necessary to have a minimum of two specific (or first rank) codes for that type of diabetes with no contradictory codes; "1" represented the presence of and "0" the absence of a definite diagnosis during each of these four years. We represented the data as two separate series of four binary numbers, the first four representing the presence or absence of a definite code for Type 2 diabetes over the last four years and the second group for Type 1 diabetes. For example, a person who had two definite codes for Type 1 diabetes might be represented by 0000–0011 and a person with two definite codes for Type 2 diabetes by 1100–0000. The possible combinations are set out in the Supporting Information.

We used a similar binary sort process to identify a probable diagnosis of a given type of diabetes. A probable diagnosis was defined when a case had at least three identical positive codes for one type of diabetes and no more than one contradictory code for the other type over the four 1-year periods. Again

we placed the four years' binary codes for Type 2 diabetes together, followed by the same four years' data for Type 1 diabetes. We sorted Type 2 diabetes data first because the prevalence of Type 2 diabetes is much more common than Type 1 diabetes. For example, 0001–1101 represents a probable diagnosis of Type 1 diabetes because there is only one Type 2 diabetes code in the 4-year period but three Type 1 diabetes codes in the second period; in contrast 1110–1000 represents a probable diagnosis of Type 2 diabetes.

We defined a possible diagnosis when there were two or more contradictory codes (e.g. 0011–1111 for possible diagnosis of Type 1 diabetes and 1111–1001 for possible diagnosis of Type 2 diabetes).

We developed a further sort mechanism for those who had diabetes diagnostic codes that did not map to either Type 1 diabetes or Type 2 diabetes; generally diagnostic labels associated with the out-of-date classification of diabetes [15] for example C108 'Insulin dependent diabetes' and C109 'Non-insulin dependent diabetes'. This was a two-step process. First, we linked less specific codes to a plausible type of diabetes (generally insulin dependent to Type 1 diabetes and non-insulin dependent or 'maturity onset' to Type 2 diabetes). We then used our filters to check whether therapy, BMI and age were compatible with this initial sort. We termed these less specific codes 'second rank' codes (Supporting Information, Table S1). We performed this search using data from any three of our four time-points; again presenting binary code for three Type 2 diabetes data points in front of those for Type 1 diabetes. A probable diagnosis would be made if there were two more identical codes in the absence of a contradictory diagnosis (i.e. 000–110; probable Type 1 diabetes or 110–000; probable Type 2 diabetes). In addition a possible Type 1 diabetes or Type 2 diabetes diagnosis would be made if three identical specific codes were recorded over the three time-points in the presence of a single contradictory code (i.e. 100–111, possible Type 1 diabetes, or 111–100, possible Type 2 diabetes).

The filters were also created using four time-points and coded in a binary format. The therapeutic filter using binary code was then generated to differentiate people on insulin, metformin or other oral anti-diabetic medication from 2005 to 2008. We classified people using the four data points identifying prescriptions of insulin, metformin or oral anti-diabetic medications one or more times in any of the years. For example, 0000–0011–0000 represents a patient not on insulin but prescribed metformin in 2007/8 in the absence of another OAD. A final filter using BMI and age was used to further differentiate between putative cases of Type 1 diabetes and Type 2 diabetes. The filter was applied at the earliest point of a diagnosis and the cut-offs for age and BMI were 35 years of age and 30kg/m² respectively.

All binary code and filter data were then extracted and recoded into SPSSTM syntax (Supporting Information, Figure S1) and a search of the database was created to classify all definite, probable and possible cases of Type 1 diabetes and Type 2 diabetes taking into account their therapeutic, anthropomorphic and biochemical data. Subjects with incomplete datasets (i.e. missing values for diagnostic codes, age, therapeutic data or biochemical results) could not be included in the analysis.

We used Bayesian statistics to test the new method compared with the previously published approach [5]. The false discovery rate [16] control, an exploratory statistical analytical method to control rigorously [17] the expected proportion of type I errors (i.e. rejected null hypothesis), was used to compare findings. The analysis was stratified into three levels: level 1 entailed false discovery rate analysis on the basis of coding alone, level 2 on the basis of coding and therapeutic data and level 3 on the basis of levels 1 and 2 and anthropomorphic and biochemical data. We then compared false discovery rates between the two search methods.

Results

The data used for this study were follow-up data ($n = 942\,031$) from the QICKD study, compared with first set of data ($n = 760,588$) that were sorted using the complex manual method.

The prevalence, in the Round 2 study data, of definite Type 1 diabetes and Type 2 diabetes were 0.32% and 3.27%, respectively (Table 1). Non-definitive (i.e. probable and possible) diagnostic codes accounted for fewer than 8% of all diagnoses of diabetes (Figure 1). Specific diagnostic coding identified 94% and 98% of Type 1 diabetes and Type 2 diabetes, respectively. Of the definite Type 1 diabetes diagnoses 31% (1004) had contradictory therapeutic data, and of those over 99% did not meet BMI/age criteria thought to be characteristic of people with Type 1 diabetes (Figure 1). Similarly, of the probable and possible Type 1 diabetes diagnoses 77% and 61% were also misclassified respectively on the basis of BMI/age criteria. Overall 35% of all Type 1 diabetes diagnoses were misclassified (Figure 1A), and nearly all should have been classified as Type 2 diabetes. Of all Type 2 diabetes diagnoses ($n = 31,438$) only 0.1% ($n = 35$) were misclassified on the basis of BMI/age at the earliest point of diagnosis (Figure 1).

We had to exclude 488 people, under 1.5% of the subjects with diabetes, from the study because we did not have enough data to sort them. This group comprised 37 with codes suggestive of steroid-induced diabetes, 46 with a possible Type 1 diabetes code; 161 with a possible Type 2 diabetes and 244 with a code that did not imply either Type 1 diabetes or Type 2 diabetes.

False discovery rate control analysis demonstrated a close correlation between the algorithmic binary search method and the independent search method in correctly identifying misclassified diabetes (Type 1 diabetes and Type 2 diabetes). At increasing levels of false discovery rate analysis for Type 1 diabetes a similar increase was noted in both the ability of the binary and independent sort methods in identifying false positives (i.e. incorrectly labelled as misclassified) (Table S1). The use of definitive diagnostic Type 1 diabetes coding alone (level 1 false discovery rate) only highlighted 5% (false discovery rate 0.05) and 3% (false discovery rate 0.03) of false positives using the binary and independent search methods respectively. The false discovery rate was highest at Level 3 (a summation of coding, therapeutic, anthropomorphic and biochemical filters for Type 1 diabetes) with 90% (false discovery rate 0.897) and 75% (false discovery rate 0.75) for the binary and independent search methods respectively.

Comparative analysis of the Level 1 false discovery rate values for Type 2 diabetes, although low overall, demonstrated a threefold difference between the manual (false discovery rate 0.06) and binary search (false discovery rate 0.02) method. Level 2 false discovery rate analysis (i.e. therapeutic coding) could not be performed, as therapy was not used in the search strategy of Type 2 diabetes.

Conclusion

A simple binary computer processable algorithm with high false discovery rate values can accurately identify misclassified and misdiagnosed cases of both Type 1 diabetes and Type 2 diabetes. The binary algorithm demonstrated combined prevalence for Type 1 diabetes and Type 2 diabetes similar to those highlighted in a recent study using the QICKD ($n = 250,000$) and CONDUIT ($n = 750,000$) data sets [5].

The binary algorithm confirms that current diabetes coding practice may overestimate the prevalence of diabetes overall and in particular Type 1 diabetes [5]. Over 30% of Type 1 diabetes were noted to be misclassified and probably have a diagnosis of Type 2 diabetes (Figure 1).

A major limitation of the study is that we could not reliably process plasma glucose data, as fasting status is often not recorded [3]. We could, however, have included plasma glucose values above 11.0 mmol/l on at least two occasions or used HbA1c ($\geq 6.5\%$, 48mmol/mol) [18]. These variables could be tested to assess their discriminatory power and then added to later refinements of the sorting tool. Any final toolkit for clinical use could also detect those with medications exclusively used to treat diabetes and hyperglycaemia. There remained a small number cases, $< 1.5\%$ of people with diabetes, who may require a manual review of their records to confirm their type of diabetes.

Problems with diagnostic diagnosis are not limited to diabetes; we have also found issues in chronic kidney disease [19], heart failure [20], osteoporosis [21] and chronic obstructive pulmonary disease [22]. Although piloted in primary care this approach could be used in any setting where diagnostic codes, therapy, age and BMI are available. This strategy could also be implemented routinely before referral to a specialist or for screening procedures, such as retinopathy screening, to check the accuracy of diagnostic data. While automated testing for coding errors is feasible, further testing is recommended before national adoption. We are close to the point where errors in the classification of diabetes might be detected automatically.

Acknowledgements

The authors thank the practices and patients for use of their anonymized data. The Health Foundation funds the QICKD study with additional support from the Edith Murphy Trust; the trial is a collaboration between Kidney Research UK and St George's – University of London. NHS Diabetes supports the Classification of Diabetes programme.

Competing interests

Both ARS and JvV have nothing to declare. KK chairs the RCGP classification of diabetes working group and had funding to conduct a systemic review of studies of misclassification. SdL has received pharmaceutical sponsorship for giving two lectures in the last two years.

References

- Schade CP, Sullivan FM, de Lusignan S, Madeley J. e-Prescribing, efficiency, quality: lessons from the computerization of UK family practice. *J Am Med Inform Assoc* 2006; **13**: 470–475.
- Alberti KG, Zimmet PZ. Definition, diagnosis and classification of diabetes mellitus and its complications. Part 1: diagnosis and classification of diabetes mellitus provisional report of a WHO consultation. *Diabet Med* 1998; **15**: 539–553.
- de Lusignan S. Flagging fasting plasma glucose specimens: time to routinely label the context in which pathology specimens are recorded. *Inform Prim Care* 2009; **17**: 63–64.
- Tai TW, Anandarajah S, Dhoul N, de Lusignan S. Variation in clinical coding lists in UK general practice: a barrier to consistent data entry? *Inform Prim Care* 2007; **15**: 143–50.
- de Lusignan S, Khunti K, Belsey J, Hattersley A, van Vlymen J, Gallagher H *et al*. A method of identifying and correcting miscoding, misclassification and misdiagnosis in diabetes: a pilot and validation study of routinely collected data. *Diabet Med* 2010; **27**: 203–209.
- de Lusignan S, Hague N, van Vlymen J, Kumarapeli P. Routinely-collected general practice data are complex, but with systematic processing can be used for quality improvement and research. *Inform Prim Care* 2006; **14**: 59–66.
- Hjerpe P, Merlo J, Ohlsson H, Bengtsson Bostrom K, Lindblad U. Validity of registration of ICD codes and prescriptions in a research database in Swedish primary care: a cross-sectional study in Skaraborg primary care database. *BMC Med Inform Decis Mak* 2010; **10**: 23.
- NHS Diabetes. Classification of Diabetes Project. Available at: http://www.diabetes.nhs.uk/our_work_areas/classification_of_diabetes/, last accessed??
- RCGP and NHS Diabetes. *Coding, Classification and Diagnosis of Diabetes. A Review of the Coding, Classification and Diagnosis of Diabetes in Primary Care in England with Recommendations for Improvement*. (DH Gateway Ref: I4664) Leicester: RCGP and NHS Diabetes, 2011.
- de Lusignan S, Gallagher H, Chan T, Thomas N, van Vlymen J, Nation M *et al*. The QICKD study protocol: a cluster randomized trial to compare quality improvement interventions to lower systolic BP in chronic kidney disease (CKD) in primary care. *Implement Sci* 2009; **4**: 39.
- de Lusignan S, Tomson C, Harris K, van Vlymen J, Gallagher H. Creatinine fluctuation has a greater effect than the formula to estimate glomerular filtration rate on the prevalence of chronic kidney disease. *Nephron Clin Pract* 2011; **117**: c213–c224.
- Bagheri A, Sadek A, Chan T, Khunti K, de Lusignan S. Using surrogate markers in primary electronic patient record systems to confirm or refute the diagnosis of diabetes. *Inform Prim Care* 2009; **17**: 121–129.
- van Vlymen J, de Lusignan S, Hague N, Chan T, Dzregah B. Ensuring the quality of aggregated general practice data: lessons from the Primary Care Data Quality Programme (PCDQ). *Stud Health Technol Inform* 2005; **116**: 1010–1015.
- van Vlymen J, de Lusignan S. A system of metadata to control the process of query, aggregating, cleaning and analysing large datasets of primary care data. *Inform Prim Care* 2005; **13**: 281–91.
- National Diabetes Data Group. Classification and diagnosis of diabetes mellitus and other categories of glucose intolerance. *Diabetes* 1979; **28**: 1039–1057.
- Benjamini Y, Hochberg Y. Controlling the false discovery rate – a practical and powerful approach to multiple testing. *J Roy Stat Soc B Met* 1995; **57**: 289–300.
- Shaffer J. Multiple Hypothesis Testing. *Ann Rev Psychol* 1995; **46**: 561–584.
- Standards of medical care in diabetes – 2010. *Diabetes Care* 2010; **33** (Suppl 1): S11–S61.
- de Lusignan S, Tomson C, Harris K, van Vlymen J, Gallagher H. Creatinine fluctuation has a greater effect than the formula to estimate glomerular filtration rate on the prevalence of chronic kidney disease. *Nephron Clin Pract* 2010; **117**: c213–c24.
- Majeed A, Williams J, de Lusignan S, Chan T. Management of heart failure in primary care after implementation of the National Service Framework for Coronary Heart Disease: a cross-sectional study. *Public Health* 2005; **119**: 105–111.
- de Lusignan S, Chan T, Wood O, Hague N, Valentin T, van Vlymen J. Quality and variability of osteoporosis data in general practice computer records: implications for disease registers. *Public Health* 2005; **119**: 771–780.
- Faulconer ER, de Lusignan S. An eight-step method for assessing diagnostic data quality in practice: chronic obstructive pulmonary disease as an exemplar. *Inform Prim Care* 2004; **12**: 243–254.

Supporting Information

Additional Supporting Information may be found in the online version of this article.

Figure S1. Extraction and binary recoding of data in SPSS.

Table S1. Binary coding for 1st rank codes and 2nd ranks codes illustrating diagnostic levels on the basis of temporal sequencing of first registered diagnosis in each year from 2005 to 2008.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

FIGURE 1

Binary algorithmic stratification of the diagnosis of diabetes using diagnostic codes and secondary filters. Classification and misclassification of Type 1 diabetes (left) and Type 2 diabetes (right) on the basis of specific and non-specific diagnostic codes with quality filtering using therapeutic, anthropomorphic and biochemical data. Data in italics are from the pragmatic independent search method.

TABLE 1

Comparative analysis of prevalence and false detection data for the binary and independent sort method. Upper: prevalence and combined sex prevalence of Type 1 diabetes and Type 2 diabetes for the chronic kidney disease dataset using the binary machine processable algorithm compared with data from the Quality Improvement in Chronic Kidney Disease (QICKD) ($n = 760\,588$). Lower: comparative false detection rate analysis using the binary and independent sort methods for Type 1 diabetes

	Binary algorithm ($N = 942,031$)	Independent sort method ($N = 760\,588$)
Prevalence		
Type 1 diabetes % (n)	0.32 (3034)	0.35 (2662)
Type 2 diabetes % (n)	3.27 (30839)	3.9 (29 663)
False detection rate analysis search filter for Type 1 diabetes		
Level 1 (definitive codes)	0.05	0.03
Level 2 (Definitive codes + therapeutic data)	0.35	0.28
Level 3 (Definitive codes + therapeutic data + anthropomorphic + biochemical data)	0.897	0.75

Figure 1: Binary algorithmic stratification of the diagnosis of diabetes using diagnostic codes and secondary filters.....