Variation in the recording of diabetes diagnostic data in primary care computer systems: implications for the quality of care

Wendy Rollason BSc
Medical Student, St George’s, University of London, UK

Kamlesh Khunti MBChB MD FRCGP
Professor of Primary Care Diabetes and Vascular Medicine, Department of Health Sciences, Leicester, UK

Simon de Lusignan MBBS MD(Res) FBCS CITP FRCGP
Reader in General Practice and Biomedical Informatics, St George’s, University of London, UK

ABSTRACT

Background Diabetes mellitus (DM) is a serious, chronic condition affecting 2.3 million people in the UK and consuming over 5% of the total National Health Service (NHS) budget. The World Health Organization (WHO) has produced a classification of diabetes which should help ensure consistent diagnosis and management of cases. However, recent quality based targets for diabetes in the UK only allow for people with Type 1 or Type 2 diabetes to be included in the disease register.

Objective To analyse the codes offered when recording a diagnosis of diabetes in an electronic patient record (EPR) system and to assess what proportion of existing codes would map to known diagnostic categories.

Method Code-sets (4-byte, 5-byte, CTv3 and SNOMED-CT) were sourced using the NHS Tri-set Browser and the SNOMED-CT website. We analysed the variation in child codes listed under ‘diabetes mellitus’. Picking lists were generated across four general practices, using eight search terms. We examined list length and the types of codes offered. An attempt was also made to map current codes to the WHO classification of diabetes, defining each as having a ‘direct mapping’, a ‘possible mapping’, or ‘no clear mapping’.

Results SNOMED-CT provided a more concise list of codes (115) than the more widely used 5-byte code-set (177). There was considerable variation in the codes offered in picking lists, with variation occurring between systems, rather than between individual GP practices. In considering the potential for mapping between current code-sets and the WHO classification, there was a general downward trend in the number that had ‘no clear mapping’ (5-byte Read codes – 46.3%, SNOMED-CT – 19.1%).

Conclusion There is considerable variation in the different diabetic coding hierarchies and in the choices offered at the point of coding in an EPR system. This is likely to lead to inconsistent data recording. Migrating GP computer systems to SNOMED-CT or to another more limited coding system which would map to international disease classifications would enable primary care EPR systems to better support improved standards of care.

Keywords: Computerised medical record systems, diabetes mellitus, systematised nomenclature of medicine

Introduction

‘Coding’ data as it is entered into computer systems makes these data much more readily searchable; and nearly all health computer systems have mechanisms for coding vital data and procedures. However, incorporating the coding process into the clinical consultation is challenging.
Revamping the Read codes would cause temporary chaos of course, but otherwise I can see no way forward for medical record keeping in primary care ... Whoever created the Read codes or, better, whoever decided to use them for GP software should be drawn and quartered.1

This quote, taken from an online news article, may echo the feelings of many with respect to the current use of Read Codes in the NHS. Remarkably, it comes from a GP over 11,000 miles away in Auckland, New Zealand; it would seem that the problem of disease coding in primary care is truly global.3

> Classification of disease is vital for effective disease management and audit of the quality of care in diabetes is no exception. Currently in the UK around 2.35 million people have been diagnosed with diabetes, and it is the leading cause of blindness in the working age population. As a highly prevalent chronic condition, with a high rate of serious long-term complications, it represents a considerable disease burden for the individual and a huge economic burden for the NHS, with over 5% of the total NHS budget spent on diabetes each year.4 Quality assurance schemes, such as the UK’s financially incentivised chronic disease management programme the ‘Quality and Outcomes Framework’ (QOF), monitor the quality of care using routinely collected computer data. This has had a direct and positive impact on patient care.5,6 However, the current diabetic QOF indicator for diabetes includes on its disease register only patients who are coded with one of two of the many possible Read codes for diabetes. The two included are: C10E and C10F and their child codes (with C10F8 Reaven’s, or metabolic syndrome, excluded).7 There are many other ways a patient might legitimately be coded with diabetes that leave them outside this particular disease register; the top level of the C10 codes (Box 1) shows the range of possible diabetes codes.8 The range of ways in which diabetes might be coded is striking and previous research has shown that recording of patients’ specific type of diabetes is poor.9 Perhaps most importantly, people coded with other diabetes mellitus (C10) codes may miss out on recall or screening linked to the QOF.

The WHO has classified diabetes and whilst its classification includes Type 1 and Type 2 the WHO recognises in addition ‘Other specific types’ of diabetes. These include: genetic defects, pancreas disease, drug- and chemical-induced causes and other rare causes. Gestational or pregnancy induced diabetes is recognised as an additional type. The WHO also recommended the ‘elimination’ of the terms ‘insulin dependent diabetes mellitus – IDDM’ and ‘non-insulin dependent diabetes – NIDDM’.10 However, whilst these terms can be removed from the WHO classification it is much harder to remove them from a classification system like the Read Codes as historic data still need to be represented within the coding system.

Given that some people with diabetes may not be captured in the current disease registers we carried out an investigation to explore the factors within the coding system and how it is put into operation in GP systems that may influence how the diagnosis of diabetes is recorded. We also wanted to explore the extent to which the common codes used could be mapped onto the WHO classification of diabetes.

### Method

#### Analysis of current diagnostic codes for diabetes

Diagnostic codes appearing under the parent code of ‘diabetes mellitus’ were obtained using the NHS Read Code Browser for the 4-byte, 5-byte and CTv3 code-sets.8 For the Systematised Nomenclature of Medicine – Clinical Terms (SNOMED-CT) the <disorder> ‘diabetes mellitus’ was located using the SNOMED-CT website, and all codes with an <is a> relationship to this were considered.11 We compared the number of codes generated for each code-set.
Variation in picking lists in different brands of GP electronic patient record system

Picking list data were copied in four general practices (from London and the surrounding area). Two were operating the EMIS LV system, one was operating the EMIS PCS and one the In Practice Systems (InPS) Vision. These brands are the most common general practice EPR systems in use in England.

Picking lists were generated as they would be during a GP consultation but without viewing any live patient data. The practices all had a simulated record they used for training. Screen shots were taken to record the picking lists generated for each search term at each practice. In order to generate representative picking lists, eight terms were used; diabetes, diab. Mel., IDDM, NIDDM, DM, Type 1, Type 2 and gest. diab.

Analysis of the data was aimed at identifying the inter-practice variation in the average length of picking lists and differences in the types of codes offered.

We specifically looked at which codes mapped to the QOF diabetes indicator disease register (i.e. C10E or C10F but excluding C10F8). We also looked for what we called ‘local’ codes, which are created by EPR system suppliers and sometimes by individual practices. Originally they were created to fill gaps in the coding system at the time. However, once created they often have to be kept in the coding system as the data they label cannot readily be mapped to a new code.

Potential mapping of diabetes codes

Having obtained a list of all diabetes diagnostic codes within each of the four code-sets, an attempt was made to map each existing code to a simplified version of the WHO diabetes diagnostic categories:

- type 1
- type 2
- gestational diabetes
- impaired glucose tolerance (IGT)
- other – this is taken to include secondary diabetes (due to other disease processes, drugs or malnutrition), some genetic forms of diabetes (sometimes termed MODY – maturity onset diabetes in the young) and other genetic causes
- undecided – refers to individuals for whom, at the time of presentation, it is difficult to define the specific type of diabetes.

Codes were deemed to have the potential for a ‘direct mapping’, a ‘potential mapping’ or ‘no clear mapping’, depending on the specificity of the current code and the degree of certainty with which it could be mapped to one of the new clinical diagnostic codes.

Results

Analysis of current code-sets

The number of diabetes diagnostic codes in each set (4-byte, 5-byte, CTv3 and SNOMED-CT) varies considerably (36, 177, 178 and 115 respectively).

The 5-byte Read code-set is the most commonly used in primary care and had a very high number of diabetes diagnostic codes. The most recently developed but less used code-set, SNOMED-CT, produced 35% fewer codes.

Picking list variation

Across the four practices there was significant variation in the mean length of the picking lists, ranging from the Vision system that generated 14 codes, to the EMIS LV systems that produced 46.9 and 50 codes. All four practices generated their longest list using ‘diabetes’ as the search term (see Table 1). The number of codes shown on each page in each system was similar, ranging from ten to 13.

The codes presented in each picking list were broken down into types of codes – diagnostic codes, history and symptom codes, examination and signs, preventative procedures and administrative. A distinction was also drawn between standard Read codes and local codes, created by the GP software and the individual practices. Mean values were calculated for

<table>
<thead>
<tr>
<th>System</th>
<th>EMIS LV #1</th>
<th>EMIS LV #2</th>
<th>EMIS PCS</th>
<th>Vision</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean list length</td>
<td>50.0</td>
<td>46.9</td>
<td>19.3</td>
<td>14.0</td>
</tr>
<tr>
<td>Number of terms on first page</td>
<td>11</td>
<td>11</td>
<td>10</td>
<td>13</td>
</tr>
<tr>
<td>Longest list length</td>
<td>111</td>
<td>139</td>
<td>30</td>
<td>68</td>
</tr>
<tr>
<td>Search term with longest list</td>
<td>diabetes</td>
<td>diabetes</td>
<td>diabetes, diab mel</td>
<td>diabetes</td>
</tr>
</tbody>
</table>
each practice to assess the types of codes presented in the set of 8 picking lists (Figure 1).

In all four practices, diagnostic codes were the most common category (mean across all practices – 49.2%), although with a mean proportion of just under half, there was clearly a wide range of other codes included. Analysis of the codes appearing on the first page gave different results, with a considerably higher proportion of them being diagnostic codes (mean across all practices – 69.83%). This could be due in part to the mechanism of ‘velocity coding’, whereby more frequently selected codes are prioritised at the top of the picking list.

In addition to the standard Read codes, all four practices also offered some local codes – i.e. codes which had been provided by the GP software, or had been created by users of the system (see Table 2).

Both EMIS LV practices offered relatively few local codes, and when considering only the first page/screen the prevalence of local codes was very low (zero in one practice). The EMIS PCS practice offered the highest number and percentage of local codes both overall and on the first screen.

Potential mapping of diabetes codes

Figure 2 shows the percentage of the number of diabetes diagnostic codes in each code-set. Considering the four code-sets as a timeline (according to their time of introduction), there has been a general downward trend in the number deemed to have no clear mapping. Comparing the widely used 5-byte set with the most recently developed set, SNOMED-CT, the percentage of codes with no clear mapping has more than halved, from 46.3% to 19.1%. The percentage of codes with a direct mapping has risen correspondingly.

As the codes used to record patient diagnoses are utilised by the QOF, it is interesting to look at how many of the QOF codes are categorised as having a direct mapping, and more importantly whether any QOF codes are considered to have no clear mapping. Both the 4-byte and 5-byte QOF codes all have a direct mapping. Three (6%) of the 50 CTV3 QOF codes have no clear mapping. Of the SNOMED-CT QOF codes, six (17%) were found to have no clear mapping and three (8.5%) to have only a possible mapping.

Discussion

Principal findings

There are several aspects of the current coding process that may lead to inconsistent data recording both between different GP systems and between individual practices using the same system:

- incoherent disease classifications

![Figure 1](https://example.com/figure1.png)

*Figure 1* Types of codes found in picking lists when searching using diabetes search terms

<table>
<thead>
<tr>
<th>Table 2 Use of local codes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Mean number of local codes in a picking list</td>
</tr>
<tr>
<td>% of local codes in a picking list</td>
</tr>
<tr>
<td>% of local codes on the first page/screen</td>
</tr>
</tbody>
</table>
Variation in the recording of diabetes diagnostic data in primary care computer systems

Numerous choices for coding of a single diagnosis combined with individual user variation

Use of local software-specific codes

Multiple methods of data entry – templates, locally optimised picking lists (velocity coding).

The analysis of the diabetes diagnostic codes available in each code-set indicated that the newer code-set, SNOMED-CT, provided a more concise list of diabetes diagnostic codes. It should, however, be noted that there were likely to be additional diabetes diagnostic codes in all code-sets that did not lie under the standard hierarchy and were therefore not included in this analysis.

Overall there seemed to be a significant difference in the length of picking lists generated between the EMIS LV systems and the Windows-style EMIS PCS and Visions systems. In general, the EMIS LV system produced much longer lists than the other two systems. In practice, however, the majority of codes selected are likely to be from the first few screens, so the absolute length of the picking list may not greatly influence the degree of diversity in codes used. In all four practices, diagnostic codes dominated the picking lists, particularly when considering those codes appearing on the first page/screen. Despite the advent of the QOF, local codes, which are by definition non-QOF codes, still featured in many picking lists, although in general they represented only a small fraction of total codes. In conclusion, there does seem to be considerable variation in the codes offered when entering a diabetes diagnosis. This variation seems to occur between different GP systems, rather than between individual practices. Analysis of the potential mapping of each of the diabetes diagnostic codes in each code-set showed an encouraging trend towards more specific and therefore more ‘mappable’ codes.

Implications of the findings

The potential for inconsistent data across GP practices could be reduced in two ways or, more likely, by a combination of both: first, the use of a code-set with fewer diagnostic codes whilst still maintaining an appropriate degree of granularity; and second, a more standardised software for entering the data.

SNOMED-CT has far fewer diabetes diagnostic codes than the more commonly used 5-byte code-set and, in addition, a higher proportion of codes that would map easily to the WHO diabetes classification. Migrating GP systems to SNOMED-CT could be considered a sensible recommendation to improve the classification of diabetes.

Standardised picking lists could help encourage standardised coding. The mapping exercise demonstrates that there is potential to flag records with ‘possible mappable’ and ‘not mappable’ diagnostic codes so that they were improved at that patient’s next visit. Codes could be classified in terms of the likelihood of the diagnosis of diabetes being definite (mappable), probable (possible mappable) and possible (no definite mapping).

Comparison with literature

Few studies have considered the quality and consistency of primary care diagnostic data. In one study of four general practices, the EPRs were compared with manual records and videoed consultations. The accuracy of the diagnostic data was found to be high. However, all four practices were self-confessed ‘high-recorders’ and all used the same GP software system. Although the data was found to be accurate, there was no measure of consistency across the four practices in the way the data was represented in the EPR. Tai et al. looked at the diversity of data entry in UK primary care and considered four GP systems – EMIS, GPASS, iSOFT and IPS. They showed that the current data entry and recording systems lead to diversity in clinical coding due to multiple factors, including velocity coding (placing commonly used codes higher in the picking list in an attempt to optimise the process of data entry) and the extensive choice of codes available to GPs when recording a single symptom or diagnosis. They concluded that a more standardised, limited list of codes...
would go some way to improving consistency in data recording.\textsuperscript{13}

Limitations of the method

This study considered only a small number of GP practices in a localised area. Although the intention was to look purely at diabetes codes, the generated picking lists also contained codes not related to diabetes. Further analysis could be undertaken, excluding these codes. This study did not assess the distribution and usage of diabetes diagnostic codes in actual patient records, merely the codes offered for use at the time of editing an EPR. A further limitation is that we restricted the mapping exercise to looking at diagnostic codes. A study carried out by QResearch found that many people who are likely to have diabetes do not have a coded diagnosis; so any systematic attempt to improve the quality of diabetic diagnostic data must also include looking for surrogate markers of diabetes, such as raised blood glucose.\textsuperscript{14}

Call for further research

Further research could look at the most appropriate level of granularity of data in primary care.\textsuperscript{15} A balance needs to be achieved whereby the choice of diagnostic codes is sufficiently limited to produce coherent data nationwide, whilst still providing an appropriate level of detail in the EPR. This would help to direct the future classification of diabetes and other conditions and would, it is hoped, provide a basis for a more stable coding system in the long term.

Once the granularity of the coding system issue is solved then efforts need to be made to see that everyone who has diabetes is mapped to the appropriate diagnostic code.

Conclusions

There are not only a huge number of codes overall to choose from when it comes to DM coding, there is also considerable variation in the codes offered by different GP systems when entering information in an EPR. Such choice and variation is likely to lead to inconsistent data.

REFERENCES


CONFLICTS OF INTEREST

None.
ADDRESS FOR CORRESPONDENCE

Simon de Lusignan
Division of Community Health Sciences
Hunter Wing
St George’s
University of London
London SW17 0RE
UK
Tel: +44 (0)20 8725 5661
Email: slusigna@sgul.ac.uk

Accepted June 2009