Refereed paper

Variation in clinical coding lists in UK general practice: a barrier to consistent data entry?

Tracy Waize Tai MBBS
Foundation Year 2 Doctor

Sobanna Anandarajah MBBS
Foundation Year 2 Doctor

Neil Dhoul BSc
Research Assistant

Simon de Lusignan MSc MD FRCGP
Reader in General Practice and Biomedical Informatics

St George’s, University of London, UK

ABSTRACT

Background Routinely collected general practice computer data are used for quality improvement; poor data quality including inconsistent coding can reduce their usefulness.

Objective To document the diversity of data entry systems currently in use in UK general practice and highlight possible implications for data quality.

Method General practice volunteers provided screen shots of the clinical coding screen they would use to code a diagnosis or problem title in the clinical consultation. The six clinical conditions examined were: depression, cystitis, type 2 diabetes mellitus, sore throat, tired all the time, and myocardial infarction. We looked at the picking lists generated for these problem titles in EMIS, IPS, GPASS and iSOFT general practice clinical computer systems, using the Triset browser as a gold standard for comparison.

Results A mean of 19.3 codes is offered in the picking list after entering a diagnosis or problem title. EMIS produced the longest picking lists and GPASS the shortest, with a mean number of choices of 35.2 and 12.7, respectively. Approximately three-quarters (73.5%) of codes are diagnoses, one-eighth (12.5%) symptom codes, and the remainder come from a range of Read chapters. There was no readily detectable consistent order in which codes were displayed. Velocity coding, whereby commonly-used codes are placed higher in the picking list, results in variation between practices even where they have the same brand of computer system.

Conclusions Current systems for clinical coding promote diversity rather than consistency of clinical coding. As the UK moves towards an integrated health IT system consistency of coding will become more important. A standardised, limited list of codes for primary care might help address this need.

Keywords: family practice, medical informatics, medical records systems – computerised, user–computer interface, vocabulary – controlled

Introduction

Using clinical concept codes consistently to represent the same disease is a well-established facet of data quality. If coding is not consistent then people with the same clinical condition may have a range of different labels. This may have implications for their clinical care if they are missed by searches; if they have
codes that fail to trigger decision support; or if incorrect use of codes means that searches contain many false positives. For example, children are all too easily given the diagnosis of ‘chronic obstructive pulmonary disease’ (COPD), a disease usually found in male smokers over 40 years old, because it appears as ‘bronchitis’ on the picking list used by clinicians looking to code acute bronchitis.\(^2\) We have also noted varying use of different coding hierarchies in mental health problems\(^3\) and ethnicity.\(^4\) The Read code includes two differing psychosis sections, one based on the International Classification of Diseases (ICD) and one that is specific to the Read classification. Ethnicity data are often recorded using multiple hierarchies so they can only be mapped into crude ethnic categories (e.g. black, Asian, white, etc.).\(^5\)

To aid consistent clinical coding, the UK has adopted a single clinical coding system. Historically there was a range of different systems but the current national standard is the Read system with the nation due to migrate to SNOMED-CT (Systematised Nomenclature of Medicine – Clinical Terms) in due course.\(^6\) The UK has taken a different approach to the rest of Europe, creating comprehensive coding systems which should represent nearly every clinical concept. By way of contrast most of our European colleagues have opted for the more compact International Classification of Primary Care (ICPC).\(^7\,8\) ICPC seeks to code the most common conditions rather than be comprehensive. Most UK practices currently use the 5-byte version of Read (version 2). Although the coding system is largely standardised it would be unhelpful if entering the same clinical term into different brands of general practitioner (GP) computer system were to reveal a different choice of Read codes.

We carried out this study to document the diversity of data entry screens available in UK general practice and its potential to influence clinical data quality.

**Method**

We carried out a literature review, using PubMed Medline. We used the search terms ‘clinical coding’, ‘classifications’, ‘vocabulary controlled’, and ‘user–computer interface’. We also contacted the GP system suppliers to identify how they selected their individual picking lists.

We identified six common conditions to represent the breadth of problem titles coded in primary care computerised medical record systems. The six problem titles we chose were: depression, cystitis, type 2 diabetes mellitus (T2DM), sore throat, myocardial infarction (MI), and tired all the time (TATT). We chose problems in different disease chapters within the Read coding system. We also deliberately selected a spectrum of problems where at one end there are distinct objective diagnostic criteria (such as diabetes) while at the other the diagnosis was the result of the clinician’s interpretation of the patient’s history (such as depression). Two conditions have a problem title that is represented as a symptom, namely ‘tired all the time’ and ‘sore throat’. We opted not to repeat diagnostic data collections in areas where we had already reported problems, namely COPD,\(^2\) psychosis,\(^3\) and ethnicity.\(^4\)

We collected screen shots of the coding picking lists that appear when typing a diagnosis into four commonly-used brands of GP computer system. The search terms (1. depression, 2. T2DM, 3. MI, 4. cystitis, 5. sore throat, and 6. TATT) were entered into whatever part of the different computer system is used to code diagnosis or the problem title of the consultation. We compared EMIS, iSOFT Synergy, IPS Vision and GPASS. We elected to use the ‘Triset’ NHS Read Code browser as the standard against which we would compare the coding alternatives suggested by the different brands of computer system.\(^9\) We documented the number of terms offered at the top level, and how many were diagnosis, symptom or other codes.

From our screen shots of the picking lists we made the following comparisons:

1. The lengths of the picking lists produced by each of the clinical systems compared with the Triset browser: These were counted based on the number that appeared on first entering the search term. If the results spread over more than one page, all the entries across all the pages were counted. Child codes for any of the choices identified were not included in the count.

2. The different types of codes appearing in each of the picking lists: We differentiated diagnostic codes (beginning with a letter); true symptom codes (beginning with a number 1); review codes (beginning with a number 8); administration codes (beginning with a number 9); the ‘R’ chapter in the Read codes, where symptoms are used as a diagnosis, we grouped with the symptom codes; and all other codes we labelled ‘other’. We also identified any system-specific or local codes, classifying them according to where they were linked into the Read code system.

3. The order of picking lists compared with the Triset browser: We also looked to see how many of the codes which appeared in each clinical system’s picking list appeared in the top ten codes of the Triset browser. We also examined whether the order codes were displayed in reflected the order in which they appeared in the Triset picking list.
Finally we looked at four practices with the same clinical system, EMIS, to see the extent to which there was inter-practice variation in picking lists. Picking lists were searched either with clinical systems logged into a training ‘patient’ or not linked to a patient record; consequently, as no identifiable patient details were identified, ethical approval was not required.

Results

We identified four mechanisms that explained why different brands of GP clinical system might order codes differently. These are presented using a taxonomy suggested by Horsfield (personal communication, 2006):

1. **Velocity coding:** This is where terms that are frequently picked are promoted to the top of the picking list.
2. **Contextual filters:** Different coding screens within the same brand of clinical computer system may employ contextual filters. For example, these offer a different range of codes if the search is for a problem title rather than a test request.
3. **Additional keying methods:** Individual brands of computer system have developed proprietary keying methods (especially for common diagnoses) different from those supplied with the codes.
4. **Multiple word entry methods:** Some brands of GP computer systems have links from partial word entries to the term (for example, ‘myo inf’ finds ‘Myocardial Infarction’).

Figure 1 Example picking lists for the Triset browser and iSOFT clinical system
The different brands of computer system produced picking lists with a mean of 19.3 codes per individual search term spread across a mean number of 2.3 screens. EMIS produced the longest picking lists (mean: 35.2 codes); GPASS produced the shortest picking lists with an average of 12.7 choices per search term. IPS also produced shorter picking lists; the average length was 13 codes per list. The Triset browser and iSOFT Synergy produced a similar mean number of codes, 17.8 and 18.8 per search, respectively. Depression had the longest picking list in all systems and was roughly double the length of the longest picking lists for that system. The Triset browser and iSOFT delivered their picking lists as a single drop-down list. GPASS and iSOFT Synergy produced single drop-down lists, unless the list was very long, as in depression. EMIS split its drop-down lists over several pages (see Table 1).

Most coding alternatives were offered for depression and fewest coding alternatives for TATT. Most of the terms offered were Read codes, though for depression there were a large number of ICD (International Classification of Diseases) codes. All of the systems had some local codes. The EMIS system consistently offered the largest number of coding alternatives across all the systems and IPS the lowest. However, each of the systems offered most choices for at least one of the clinical conditions entered.

Overall approximately three-quarters (73.5%) of the codes offered were diagnostic terms; symptom codes accounted for about one-eighth of the codes (12.5%); review and administration codes only appeared in significant numbers for depression; and the remainder (8.7%) were codes from ‘other’ chapters (see Table 2). Depression was a special case: in addition to returning large numbers of diagnostic terms and producing long lists of review and administrative codes, its diagnostic codes were of two types. Depress- sion codes either begin with an ‘E’ or an ‘Eu’. The Eu codes are designated on picking lists by an [X] in front of the code. [X] signifies inclusion in the ‘Behavioural and mental health disorder’ chapter. They represent ICD10 (World Health Organization International Classification of Diseases version 10) codes, which are not mapped to Read codes. They form a parallel hierarchy within the ‘E’ chapter. Although the Eu codes are predominantly used in secondary care, they have differing prominence across GP computer systems. Thirty-eight percent of the Triset codes for depression are Eu codes, a larger proportion than that produced by any of the other picking lists. The mean proportion of Eu codes was 31%, with two systems, GPASS and IPS, displaying above the average number and two systems displaying below the average. However, as explained in more detail below, GPASS places the Eu codes high up in its picking list.

There was little commonality in the order in which the codes were offered, either by the individual clinical systems or when compared with the Triset browser; we could find no identifiable pattern. To try to quantify differences in the order, we looked to find where in the picking list of each different brand of computer system the top ten items on the Triset browser appeared. No relationship could be identified for any of the variables other than when the list was very short. Data for depression and MI are shown in Table 3. Although IPS shared nine codes with the Triset top ten, this pattern was not consistently repeated across all six problem titles. The ranking of depression codes in GPASS was particularly interesting; 11 out of the first 13 codes (actually all but the first two) were ‘Eu’ – ICD10 mapped codes rather than Read codes. No similar obvious code groupings were apparent for the other problem titles.

### Table 1 Number of items on picking lists and their length in pages

<table>
<thead>
<tr>
<th></th>
<th>Triset</th>
<th>EMIS</th>
<th>GPASS</th>
<th>IPS</th>
<th>iSOFT</th>
<th>Overall mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean number of codes</td>
<td>17.8</td>
<td>35.2</td>
<td>12.7</td>
<td>13.0</td>
<td>18.8</td>
<td>19.5 codes</td>
</tr>
<tr>
<td>per picking list</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shortest picking list</td>
<td>1</td>
<td>8</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>3.4 codes</td>
</tr>
<tr>
<td>Search term for shortest list</td>
<td>TATT</td>
<td>TATT</td>
<td>TATT</td>
<td>Sore throat</td>
<td>Type 2 DM</td>
<td></td>
</tr>
<tr>
<td>Longest picking list</td>
<td>57</td>
<td>96</td>
<td>50</td>
<td>50</td>
<td>81</td>
<td>66.8 codes</td>
</tr>
<tr>
<td>Search term for longest list</td>
<td>Depression</td>
<td>Depression</td>
<td>Depression</td>
<td>Depression</td>
<td>Depression</td>
<td></td>
</tr>
<tr>
<td>Mean number of pages</td>
<td>1</td>
<td>4</td>
<td>1.8</td>
<td>1.8</td>
<td>3.0</td>
<td>2.3 pages</td>
</tr>
<tr>
<td>of codes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Table 2 Types of Read codes displayed

<table>
<thead>
<tr>
<th></th>
<th>Depression</th>
<th>Type2 DM</th>
<th>MI</th>
<th>Cystitis</th>
<th>Sore throat</th>
<th>TATT</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n (% )</td>
<td>n ( % )</td>
<td>n ( % )</td>
<td>n ( % )</td>
<td>n ( % )</td>
<td>n ( % )</td>
<td>n ( % )</td>
</tr>
<tr>
<td>No. of diagnostic terms</td>
<td>236 (70.7)</td>
<td>48 (100)</td>
<td>72 (72.7)</td>
<td>43 (87.8)</td>
<td>23 (71.9)</td>
<td>8 (34.8)</td>
<td>430 (73.5)</td>
</tr>
<tr>
<td>No. of symptom terms</td>
<td>31 (9.3)</td>
<td>0 (0)</td>
<td>21 (21.2)</td>
<td>6 (12.2)</td>
<td>9 (28.1)</td>
<td>6 (26.1)</td>
<td>73 (12.5)</td>
</tr>
<tr>
<td>Review codes</td>
<td>12 (3.6)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>12 (2.1)</td>
</tr>
<tr>
<td>Administrative codes</td>
<td>19 (5.7)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>19 (3.2)</td>
</tr>
<tr>
<td>No. of other terms</td>
<td>36 (10.8)</td>
<td>0 (0)</td>
<td>6 (6.1)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>9 (39.1)</td>
<td>51 (8.7)</td>
</tr>
<tr>
<td>Total</td>
<td>334 (100)</td>
<td>48 (100)</td>
<td>99 (100)</td>
<td>49 (100)</td>
<td>32 (100)</td>
<td>23 (100)</td>
<td>585 (100)</td>
</tr>
</tbody>
</table>

## Table 3 Where the top ten Triset codes for myocardial infarction and depression sit in the picking lists of five clinical systems

### Myocardial infarction (MI)

<table>
<thead>
<tr>
<th>Triset codes</th>
<th>Row where code was found</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>EMIS</td>
</tr>
<tr>
<td>1 Old myocardial infarction</td>
<td>34</td>
</tr>
<tr>
<td>2 FH: Myocardial infarction</td>
<td>12C3. NP</td>
</tr>
<tr>
<td>3 ECG: myocardial infarction</td>
<td>323..</td>
</tr>
<tr>
<td>4 Acute myocardial infarction</td>
<td>G30..</td>
</tr>
<tr>
<td>5 Myocardial infarction aborted</td>
<td>G3110</td>
</tr>
<tr>
<td>6 FH: Myocardial infarction &gt;60</td>
<td>12C3.</td>
</tr>
<tr>
<td>7 FH: Myocardial infarction &lt;60</td>
<td>12C2.</td>
</tr>
<tr>
<td>8 Subsequent myocardial infarction</td>
<td>G35..</td>
</tr>
<tr>
<td>9 Postoperative MI</td>
<td>G38..</td>
</tr>
<tr>
<td>10 [V]Observation for suspected MI</td>
<td>ZV719</td>
</tr>
</tbody>
</table>

Number of MI codes in Triset top ten: 585 (100)

continued
We looked at the picking lists of four EMIS computer systems to see whether ‘velocity coding’, the process whereby commonly-used codes are accelerated to the top of the picking list, resulted in very different picking lists. EMIS was used as a convenience sample. Again there appeared to be some commonality between the top 10–15 codes, but considerable variation in order. The results for MI are shown in Figure 2. All the picking lists start with G30 ‘Acute myocardial infarction’, and two out of four have G32 ‘Old myocardial infarction’ as the second code. Codes G308 ‘Inferior myocardial infarction’ and 12C ‘Family history’ are common to all the practice picking lists. There is considerable variation between the other codes used.

**Discussion**

The clinical coding picking lists are not consistent in how they browse the same underlying coding system, with the result that they might promote diverse coding practice. This variation in length, type of code, and order, is likely to undermine rather than stimulate improvement in data quality. ‘Velocity coding’ may make inter-practice variation in choice of codes more marked, as the more a practice uses a code the more prominent it will be within its picking list. This may be acceptable if coding is a result of careful informed choice about what codes to use, but is potentially damaging if poor coding choices are made. The double structure of the mental health or ‘E’ chapter is unhelpful; clinical systems should consistently use either the Read or ICD10 hierarchy, not both.

There may be a case for a shorter general practice code list. The wisdom of having such a large number of coding choices within Read version 2 for common conditions needs careful review. Whilst it is attractive that there should be a Read code for every clinical concept, the cost of this is long picking lists that clinicians may not have time to consider fully; a shorter common GP list could result in more consistent coding and improved data quality. Patients are generally booked into ten-minute GP consultations, and the time taken to code the diagnosis or problem presented needs to be kept to a minimum.

Inter-practice variation in clinical coding has been recognised as a problem for some time.11 The variation in picking lists probably contributes towards this. This is a practical problem when aggregating data from different practices. We have demonstrated this for ethnicity data, where different practices use differing coding hierarchies.4
Research is needed to know whether a more limited list of codes might improve data quality and shorten the time taken to code. The codes nominated for inclusion in the Quality and Outcomes Framework (QOF) could be one sensible starting point, the International Classification of Primary Care (ICPC) another. QOF codes are now well embedded into chronic disease management in UK primary care and it might damage patient care to migrate away from them. ICPC offers advantages in that it is a limited code list, is widely used in Europe (thereby facilitating patient record flows), and it has been extensively mapped to ICD10 and other coding systems. Both in the UK and internationally there have been proposals for a limited GP vocabulary within SNOMED-CT.
Conclusions

The major clinical systems all have very different interfaces with the same underlying coding system. Entering the same search term in different brands of computer system led to different numbers of varying types of code being presented to the user. Increasing mobility of the population, the introduction of GP-to-GP electronic transmission of records, and the move towards an integrated NHS IT system, makes the consistent use of clinical codes more important. Standardising the picking lists generated by the different brands of computer system is desirable and would remove a potential barrier to consistent clinical coding.

ACKNOWLEDGEMENTS

Helpful comments and a supply of screen shots were received from: Pete Horsfield, Frank Sullivan, Nigel Hague and John Williams. We thank Vickie Wood for her careful correction of the manuscript.

REFERENCES

1 Pringle M, Ward P and Chilvers C. Assessment of the completeness and accuracy of computer medical records in four practices committed to recording data on computer. British Journal of General Practice 1995;45:537–41.
5 Primary care electronic library (PCEL). Ethnicity Recording in General Practice Computer Systems. 3. Mapping of the 95 and 96 ethnic codes to the ‘16+1’ format and the five category classifications. www.pcel.info/ethnicity/
14 University of Adelaide, Faculty of Health Sciences. General Practice Vocabulary Project.

CONFLICTS OF INTEREST

None.

ADDRESS FOR CORRESPONDENCE

Simon de Lusignan
Senior Lecturer, Primary Care Informatics
St. George’s, University of London
Hunter Wing
London SW17 0RE
UK
Tel: +44 (0)20 8725 5661
Fax: +44 (0)20 8725 3584
Email: slusigna@sgul.ac.uk

Accepted May 2007