Conference report

The optimum granularity for coding diagnostic data in primary care: report of a workshop of the EFMI Primary Care Informatics Working Group at MIE 2005

Simon de Lusignan MSc MD MRCGP
Chair, EFMI Primary Care Informatics Working Group, Primary Care Informatics, Division of Community Health Sciences, St George’s – University of London, London, UK

ABSTRACT

Introduction The EFMI Primary Care Informatics Working Group held a workshop to explore interventions used across Europe to improve the data quality in primary care computerised medical records.

Method A plenary session reviewed the UK literature about improving data quality and then the session split into three small groups. Fifteen delegates from nine countries contributed to the workshop. These groups reported back at the end of the session.

Results The groups defined what they meant by data quality. The principal requirement was that data must be ‘fit for purpose’. The participants felt this was particularly important for diagnostic data, while recognising that the purpose might not be known at the point of data recording. They also described the barriers to recording structured and coded data. The most important were an inappropriate interface with the coding system and inappropriate granularity of codes. There was a wide range of suggestions as to how to overcome these barriers, including providing feedback, links to expert systems, education and training, use of the data for care elsewhere in the health system and mandation of electronic data recording.

Conclusions The workshop developed a new characteristic of data quality: ‘fit for purpose’. This is different from definitions that focus on completeness, accuracy, currency, or its positive predictive value and sensitivity. The group also highlighted the importance of data quality of diagnoses, as these data are important throughout the health system as well as acting as a prompt for other interventions within the individual consultation. More research is needed into appropriate levels of granularity for diagnostic recording in primary care.

Keywords: clinical data quality, coding diagnostic data, primary care computer records

Introduction

Primary care computer records should contain good quality data. However, what defines data quality and what interventions promote high-quality data remains open to debate. Early definitions of data quality focused on the completeness and accuracy of the data. Williams suggested that currency (how recently the data were recorded) should be added. In a systematic review about the quality of general practice data, Thiru et al suggested that the positive predictive value (that is, the likelihood that someone with a diagnostic label actually had the diagnosis) and sensitivity (that is, the number of false negatives) were the most important measures of data quality. A range of interventions have been proposed to improve data quality: in the UK feedback, education and training have been the mainstay of interventions to improve data quality, though other methods, such as data quality probes, have also been used.
The Primary Care Informatics Working Group (PCI WG) is one of the largest and most active in the European Federation for Medical Informatics. The PCI WG aims to support the development of the science and practice of primary care informatics. The group’s four domains of interest include promoting data quality in primary care records, a theme we have also developed in collaboration with our primary care colleagues in the International and American Medical Informatics Associations (IMIA and AMIA).

This workshop was designed to explore whether there was a consensus across EFMI countries as to what data quality is and how best to achieve it in primary care.

**Method**

The workshop was included as part of the programme of the Medical Informatics Europe (MIE) 2005 conference in Geneva in August 2005. It was planned with the following objectives:

1. to describe interventions used across Europe to improve clinical data quality in primary care computer records
2. to identify common principles and practice in interventions designed to improve data quality.

The workshop was planned to consist of a short presentation of background material, small-group sessions and a final plenary session. The presentation was based on a literature review of the barriers to recording clinical data. The conclusion of this review was that there are four types of barriers to recording structured coded data and that this analysis might form a useful taxonomy for the workshop. The four headings were:

1. Individual factors: the individual’s knowledge, skill and motivation to record clinical data
2. Consultation factors: how to integrate coding into the clinical consultation and the effect of using predominantly biomedical labels in primary care
3. Technological factors: the interface with the computer and the coding system
4. Organisational factors: influences at the levels of the practice, locality, region and health service.

Small-group discussions followed the presentation, designed to identify common features of successful data quality programmes and including an opportunity for delegates briefly to present their experience in using feedback, or other interventions, to improve data quality in primary care.

The final session was a plenary presentation of the small groups’ findings and an attempt to derive a consensus statement. The output from the workshop was then circulated to delegates for their comments.

Notes taken during the workshop, flip charts and the presentation provided the material used to document the event. One UK delegate kept notes throughout in a field notebook. One delegate in each small group agreed to keep detailed notes of their small group. These notes and flip charts were transcribed afterwards and used as the basis for this report.

**Results**

**Delegates and the coding systems used**

A broad range of European countries were represented at the workshop. Fifteen delegates from nine countries attended the workshop, of whom 13 left functioning email addresses. The countries represented were: UK (four delegates), Romania (two), Turkey (two) and one delegate each from Belgium, Iceland, The Netherlands, Serbia, Sweden and Yugoslavia. The delegates’ countries used a range of clinical coding systems (see Table 1).

**Defining data quality**

The first group task was to agree a definition of data quality; the groups reiterated the existing definitions but added new features, particularly that data must be ‘fit for purpose’. The existing definitions confirmed were that data should be complete, accurate and current. In addition, the groups added that data should be relevant (to the person’s health care) and accessible. This echoes the English data quality programme ‘PRIMIS+’ definition of data quality: ‘CARAT’ (Complete, Accurate, Relevant, Accessible, Timely), see www.primis.nhs.uk. By ‘accessible’, the groups meant that the individual patient’s data could be accessed when needed for their care and that a population’s data were accessible for health service planning and other secondary purposes. One group preferred data quality to be defined as whether the data were ‘fit for purpose’, meaning that completeness, accuracy, currency, relevance and accessibility were not ends in themselves: what mattered was whether data were suitable for their intended purpose.

**Barriers to recording coded data**

Many barriers to clinical coding were identified, which can be mapped to the proposed taxonomy. The difficulty in ‘restraining GPs from doing their own thing’ was seen as the principal barrier at the individual clinician level. Coding was seen as an ‘overhead’ in
The optimum granularity for coding diagnostic data in primary care

The clinical process. There seemed agreement in all the groups that primary care clinicians don’t like ‘coding’, especially during the consultation: they neither like the time taken to code nor the process of attaching labels to vague problems. Technical problems were also described when selecting codes from picking lists.

Issues of confidentiality and consent were also considered important barriers to recording coded data and delegates felt these issues needed technical solutions. Patients should give their consent to allowing their data to be used even in ‘anonymised’ databases. De-identified data, with strong identifiers removed, might still not be anonymous. This is especially so for rare diseases (for example, hyperparathyroidism) and where large numbers of variables are included in a dataset.

Overcoming the barriers to coding clinical data

No single intervention was identified that would overcome the barriers to recording coded clinical data. Training and provision of feedback were reiterated as ways of providing individual clinicians with motivation to code clinical data; in addition new mechanisms were added to this list. Recording high-quality diagnostic data might trigger appropriate decision support, flagging up guidelines for improved management; it might also provide data that can be used for health service planning and research. In the Netherlands, diagnostic data have been linked to expert systems to promote coding.12

Integration of systems so that coded data recorded in one part of the health system were available for

<table>
<thead>
<tr>
<th>Table 1 Clinical coding systems and number of primary care vendors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Country</td>
</tr>
<tr>
<td>UK</td>
</tr>
<tr>
<td>Belgium</td>
</tr>
<tr>
<td>Romania</td>
</tr>
<tr>
<td>Turkey</td>
</tr>
<tr>
<td>Sweden</td>
</tr>
<tr>
<td>The Netherlands</td>
</tr>
<tr>
<td>Serbia</td>
</tr>
<tr>
<td>Iceland</td>
</tr>
</tbody>
</table>

ICPC, International Classification of Primary Care; ICD, International Classification of Disease; NIC, Nursing Interventions Classification.
Patient care in another was cited as a very important reason to code diagnostic and other data. In The Netherlands, coded data are available to the out-of-hours emergency service for patient care. If a general practitioner (GP) does not record coded data then important facets of a patient’s clinical record will not be available for the hours in the day when the surgery is closed. This has motivated GPs in The Netherlands to improve their coding of data.

It was suggested that the way in which an individual clinical computer system interfaced with the coding system might bias the selection of codes: the more complex the coding system, the more important this was likely to be; for example, different vendors’ coding interfaces listed codes in varying orders when the same letters were entered (such as DIAB when trying to find appropriate codes for diabetes diagnostic or diabetes-related codes). GPs rarely spent much time in a consultation looking for the right code; instead they were likely to select an option appearing near the top of the picking list.

Coding systems were often too detailed for the sophistication of diagnosis possible in primary care. The group discussed the diagnosis of pneumonia as an exemplar. A GP might be able to diagnose clinically that a patient has pneumonia, but nothing more sophisticated than that, unless he has completed tests or is coding an episode of hospital care. The International Classification of Primary Care (ICPC) appears to offer a level of granularity appropriate to the level of sophistication of clinical coding that a GP might want to make. The coding choices in a Netherlands system are bronchopneumonia or other pneumonia. This simple classification obviates the need to look through more complex codes. The number of alternatives found when looking to code pneumonia in ICPC, in the World Health Organization’s International Classification of Disease (ICD) and Read Clinical Terms version 3 (CTv3), are shown in Box 1.

Clear national policy and possibly strict mandates can only help coding. A financially incentivised, quality-based contract, with achievement only based on computer coding, has led to increased coding of diagnosis and other clinical data within the UK.13

**Discussion**

The principal findings of the workshop were to redefine data quality in functional terms and to encourage those interested in clinical coding to think about the importance of having a granularity appropriate to the clinical task.

Data quality has previously been described in the literature in mathematical terms (completeness, accuracy, currency, positive predictive value and sensitivity), all of which can be measured numerically. This workshop concluded that a functional approach – are these data fit for the purpose intended? – is a more useful approach.

No recent research is reported in the medical literature about the optimum granularity for a primary care coding system. This appears to be a very important question and one worth exploring on an international basis. There appears to be face validity in the assertion that GPs faced with overlong picking lists have their choices biased by what is near the top of the picking list. There appeared consensus that these long picking lists

---

**Box 1** Coding ‘pneumonia’ using ICPC, ICD-10 and Read CTv3

1 **Coding pneumonia with ICPC (International Classification of Primary Care)**
   A GP in The Netherlands using ICPC is presented with a choice of two options:
   (i) Bronchopneumonia OR
   (ii) Other pneumonia.
   If this GP wishes to code in more detail they must use ICD.

2 **Coding pneumonia with ICD-10 (International Classification of Diseases and Health-related Problems)**
   Icelandic GPs code problems using this system. They are presented with a choice of 80 codes. These relate to different causes of pneumonia including congenital cause, infections that may result in pneumonia, and the type.

3 **Coding pneumonia with CTv3 (Read Clinical Terms version 3)**
   A UK GP using this system will be presented with a choice of 182 alternatives. These alternatives have finer granularity than those offered by ICD, e.g. **Right middle zone pneumonia**. Many of these have lower levels with further codes available.
lists often had a level of granularity inappropriate for primary care.

To test this out we have decided to focus a future workshop on examining the picking lists presented to GPs across Europe when they set out to code some reference conditions:

1. myocardial infarction
2. type 2 diabetes mellitus
3. depression
4. tired all the time (TATT – with no evidence of organic disease or depression)
5. sore throat (not tonsillitis)
6. cystitis (frequency, dysuria but urine not tested).

Footnote (1 April 2006)

Working Group members were contacted by email to see if we could complete this work for the Special Topic Conference 2006. However, due to pressure of work many members of the Working Group were not able to respond. We have data from the different English primary care clinical computer systems and from the principal systems used in Scotland, but not as yet from other EFMI countries. We look forward to developing this area of interest over successive Working Group meetings.

Important future events for the Primary Care Informatics Working Group

A Primary Care Workshop was held at MIE 2006 (26 August 2006, Maastricht, The Netherlands): The workshop title was: Routinely collected GP data: Goldmines for Research. The output from this workshop, attended by 23 delegates, will be submitted for publication in this journal. The output included 10 Maastricht rules which should be applied prior to using routinely collected GP data for research.

STC 2007 Croatia: A primary care day has been organised as part of the EFMI STC (European Federation of Medical Informatics Special Topic Conference) in Croatia 28–30 May.

Primary Care Day at MEDINFO 2007: On Sunday, 19 August 2007, there will be a one-day primary care consensus conference prior to the start of MEDINFO, in Brisbane, Australia. This will be jointly organised with PCI WGs from IMIA and AMIA.

REFERENCES

12. ‘The transition project. www.transitieproject.nl/

CONFLICTS OF INTEREST

None.

ADDRESS FOR CORRESPONDENCE

Simon de Lusignan
Chair, EFMI Primary Care Informatics Working Group
Primary Care Informatics
Division of Community Health Sciences
St George’s – University of London
London SW17 0RE
UK
Tel: +44 (0)208 725 5661
Fax: +44 (0)208 767 7697
Email: slusigna@sgul.ac.uk

Accepted April 2006