Conference papers

A methodology for the functional comparison of coding schemes in primary care

Philip JB Brown MD MRCGP
School of Information Systems, University of East Anglia, Norwich, and Humbleyard Practice, Hethersett, Norfolk, UK

Victoria Warmington RGN BSc
Humbleyard Practice, Hethersett, Norfolk, UK

Michael Laurence MA MRCGP
Bacon Road Medical Centre, Norwich, Norfolk, UK

A Toby Prevost MSc PhD
Department of Public Health and Primary Care, University of Cambridge and Institute of Public Health, Cambridge, UK

ABSTRACT

There has been massive investment in the development of clinical terminologies for use in electronic patient records. However, there has been little published evidence for the added value for primary care that implementation of such a terminology would offer. This paper outlines a methodology that has been used to compare two existing coding schemes (Read codes 5 byte set and Clinical Terms Version 3 – CTV3) and demonstrates their relative performance using a certainty–agreement diagram. In the study described, CTV3 offers improved accuracy and consistency with improved usability. The potential advantages of the recently released terminology, SNOMED Clinical Terms, are briefly considered in this context.

Keywords: clinical terminologies, coding schemes, electronic patient records

Introduction

The National Health Service (NHS) over the last decade has invested millions of pounds in the development of formally structured electronic coded vocabularies (clinical terminologies). The development of Clinical Terms Version 3 Read codes (CTV3) during the Clinical Terms Projects culminated in its release in 1994.[1,2] However, uptake in primary care has been poor with only a few developers, such as Healthy Software and Phoenix, investing in the design of systems. Various reasons have been mooted for this, including complexity of structure of CTV3 and a lack of interest amongst users.

The situation was complicated further in 1999 with the announcement of the merger between the United Kingdom (UK) CTV3 and the United States (US) terminology, SNOMED-RT, developed by the College of American Pathologists (CAP). The product of two years of collaborative effort between the NHS Information Authority and CAP is SNOMED Clinical Terms (SNOMED-CT).[3] This has been released in the UK this year for testing and is an important proposed component of the NHS Information Strategy, and the US Government has recently signed a nationwide licence. Thus SNOMED-CT – a clinical terminology with structure of similar complexity and design as CTV3 – is poised to become a proto-global standard.

It is therefore surprising that, despite the enormous investment of effort and resources, there are few formal comparative evaluations between different coding schemes to establish that such ventures will improve electronic records and ultimately patient care. The reality of existing CTV3 systems proves that implementation of a clinical terminology in primary care is possible. For users and system developers to adopt a clinical terminology (such as SNOMED-CT), evidence is needed of the practical advantages that will accrue to compensate the costs of migrating from the coding schemes in current use (Read codes 5 byte set [RC5B] and Read codes 4 byte set).

In order to address this void in evidence we have recently described a methodology of a randomised crossover trial comparing the performance of CTV3 and RC5B coding schemes in general practice.[4] This study looked at the comparative performance of the two coding schemes by ten general practitioners (GPs) using a total of 995 concepts extracted from clinical records.

In this paper we briefly outline the methodology used and describe the novel use of the certainty–agreement diagram to represent the differences graphically as a practical method of demonstrating the potential advantages of using a clinical terminology over existing coding schemes in general practice.

Methodology

The RC5B is widely used in UK general practice. The study sought to compare the performance of this existing scheme with a clinical terminology (CTV3). The methodology has been described elsewhere, but the salient features are summarised here for convenience.[4]

Term collection

Ten GPs collected a corpus of terms from a mixture of urban, rural and semi-rural environments. These terms were collected manually from ten consecutive consultations (a total of 100) using a paper proforma encouraging the listing of relevant terms under standard headings; for example, reason for encounter, past medical history and so on. This process was designed purely as a collection exercise for terms relevant to primary care consultations and was not meant to reflect the actual process of data entry in the 'live' situation. For example, during a consultation one might enter the reason for encounter or the main diagnosis(es) but would rarely encode the complete patient past medical history.

Encoding exercise

The corpus of terms was then entered verbatim into an Access™ database and duplicates removed. The resultant 995 unique concepts were then used as a valid source of primary care-derived terms. The ten GPs were randomly allocated into pairs and each pair was given approximately 200 concepts to encode. One GP of the pair was asked to encode the set using the CTV3 coding scheme first followed by RC5B, the matched paired GP was asked to encode the same set using RC5B first followed by CTV3. All GPs used the same hardware (Sony Vaio) and software (NHS IA Clinical Terminology Browser) in order to minimise software and hardware confounding variables. Each GP's coding activity was videoed and they were asked to narrate their coding activity in a similar way to that employed by Cimino *et al*.[5]

Performance measures

The performance of each coding scheme was assessed by the following measures:

- *Exactness* of representation was assessed independently by two clinicians who examined the original concept and the choice of coded term. A judgement was made as to whether the match was exact or non-exact. Conflicts of agreement between the two clinicians were resolved by consensus agreement.
- *Agreement* was assessed by identifying the proportion of concepts that were matched to the same code by each of the paired GPs.
- *Usability* was measured by reviewing the video of the coding activity and recording the time taken to code the concept and the number of attempts to find an appropriate match (that is, the number of times the user had to re-key possible terms into the software).

Certainty–agreement diagram

This paper describes the use of the certainty–agreement diagram in providing a graphical representation of comparative functionality. Biological and social systems are inherently complex, and by
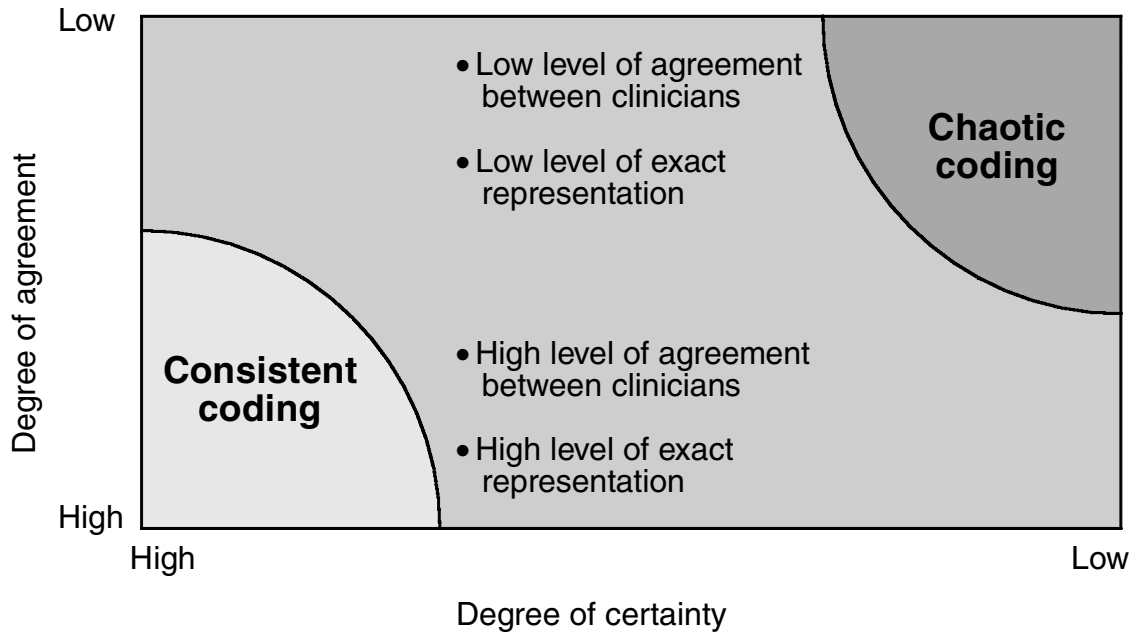
Figure 1 Certainty–agreement diagram (adapted from Wilson, Holt and Greenhalgh[6])[7,8]

implication so are the coding schemes that have to capture these data. The certainty–agreement diagram (Figure 1) can be used to estimate whether the issue is simple (high certainty, high agreement), chaotic (low certainty, low agreement) or complex (intermediate levels of one or both).[6]

Results

The performance of CTV3 and RC5B is graphically illustrated in Figure 2, which illustrates that there is significant improvement in agreement and certainty of a GP's ability to code using CTV3 compared to RC5B. In an ideal coding scheme there should be 100% success in exactly representing the concept one requires using the scheme and the choice should be perfectly consistent between users. If CTV3 were perfect in this respect the plots would group around the intersect of the x and y axes.

From previously published results, the difference in performance between CTV3 and RC5B has been shown to be significant with exact matches in CTV3 being significantly more common (70% [95% CI 67–73]) than with RC5B (50% [47–53]), $P<0.001$, and this was statistically significant for each of the ten GPs individually. The pooled proportion with exact and identical matches by paired GPs was greater for CTV3 (0.58 [0.55–0.61]) than RC5B (0.36 [0.33–0.39]), $P<0.001$.[4]

Concerns might exist that the improvement in accuracy of representation (exactness) and consistency (agreement) might have been achieved at the expense of usability. However, the time taken to code using CTV3 (30.2 seconds, 95% CI 28.6–31.9) was shorter than that for RC5B (36.1 seconds, 95% CI 34.3–37.9).[4] The incidence of re-keying attempts per concept was also less for CTV3 (0.39) compared to RC5B (0.45).
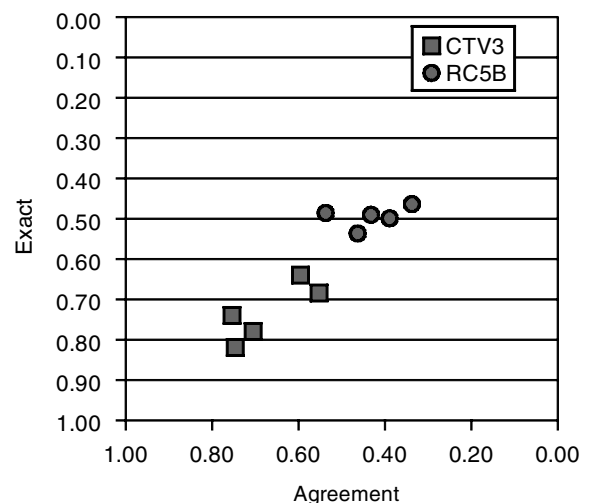

Figure 2 Certainty (proportion exact)–agreement diagram of coding between matched doctors using CTV3 and RC5B

Discussion

The methodology has attempted to minimise confounding variables to focus on the comparative performance of two coding schemes. The full functionality of a terminology includes the ability to add further detail by the use of qualifying a core concept with additional detail. This would have unfairly biased the result towards CTV3 and this functionality was therefore not tested in this study. However, if this mechanism were employed, the measure of accuracy (exactness) would be expected to increase – but it is uncertain as to whether given even greater choice the agreement between users might suffer.

The results are encouraging evidence that the use of a terminology in primary care information systems will improve the accuracy and consistency of data capture. As SNOMED-CT incorporates and has an enhanced content in comparison with CTV3, one would anticipate that it should offer similar advantages; however, confirmation is needed that its greater coverage (exactness) is not achieved at the expense of deterioration in measures of consistency and usability.

The complete and consistent collection of data within general practice is a complex process and is affected by many factors other than the coding scheme, including software factors, time constraints, user training and so on. If the laudable objectives of accurate data collection and retrieval are technically possible, investment in system design, training and operational management (for instance, who collects the data, allocated time) needs addressing both in terms of further research and workable practical solutions.[9]

REFERENCES

1 Stuart-Buttle CDG, Brown PJB, Price C, O'Neil MJ and Read JD. The Read Thesaurus – creation and beyond. In: Pappas C, Maglaveras N and Scherrer J-R (eds) *Proceedings of the 14th Medical Informatics Europe '97 Conference.* Oxford: IOS Press, 1997:416–20.

2 O'Neil M, Payne C and Read JD. Read Codes Version 3 – a user-led terminology. *Methods of Information in Medicine* 1995;34:187–92.

3 Stearns MQ, Price C, Spackman KA and Wang AY. SNOMED Clinical Terms: overview of the development process and project status. In: Bakken S (ed) *Proceedings of the 2001 AMIA Fall Symposium.* Philadelphia: Hanley and Belfus, 2001:662–6.

4 Brown PJB, Warmington V, Laurence M and Provost AT (2003) Randomised crossover trial comparing the performance of Clinical Terms Version 3 and Read codes 5 byte set coding schemes in general practice. *British Medical Journal* 2003;326:1127–30.

5 Cimino JJ, Patel VL and Kushniruk AW. Studying the human–computer terminology interface. *Journal of the American Medical Informatics Association* 2001;8:163–73.

6 Wilson T, Holt T and Greenhalgh T. Complexity science: complexity and clinical care. *British Medical Journal* 2001; 323:685–8.

7 Stacey RD. *Strategic Management and Organisational Dynamics: the challenge of complexity* (3e). London: Financial Times, 1999.

8 Zimmerman B, Lindberg B and Plsek P. *Edgeware: insights from complexity science for health care leaders.* Irving, TX: VHA Press, 1998.

9 Brown PJB and Sönksen P. Evaluation of the quality of information retrieval of clinical findings from a computerised patient database using a semantic terminological model. *Journal of the American Medical Informatics Association* 2000;7:401–12.

FUNDING

ADDRESS FOR CORRESPONDENCE

Dr Philip JB Brown
School of Information Systems
University of East Anglia
Norwich NR4 7TJ
UK
Tel: +44 (0)1603 810250
Fax: +44 (0)1603 812402
Email: phil@hicomm.demon.co.uk