Information in practice

Randomised crossover trial comparing the performance of Clinical Terms Version 3 and Read Codes 5 byte set coding schemes in general practice

Philip J B Brown, Victoria Warmington, Michael Laurence, A Toby Prevost

ELPS

This is an abridged version; the full version is on bmj.com

Abstract

Objective To determine whether Clinical Terms Version 3 provides greater accuracy and consistency in coding electronic patient records than the Read Codes 5 byte set.

Design Randomised crossover trial. Clinicians coded patient records using both schemes after being randomised in pairs to use one scheme before the other.

Setting 10 general practices in urban, suburban, and rural environments in Norfolk.

Participants 10 general practitioners.

Source of data Concepts were collected from records of 100 patient encounters.

Main outcome measures Percentage of coded choices ranked as being exact representations of the original terms; percentage of cases where coding choice of paired general practitioners was identical; length of time taken to find a code.

Results A total of 995 unique concepts were collected. Exact matches were more common with Clinical Terms (70% (95% confidence interval 67% to 73%)) than with Read Codes (50% (47% to 53%)) ($P < 0.001$), and this difference was significant for each of the 10 participants individually. The pooled proportion with exact and identical matches by paired participants was greater for Clinical Terms (0.58 (0.55 to 0.61)) than Read Codes (0.36 (0.33 to 0.39)) ($P < 0.001$). The time taken to code with Clinical Terms (30 seconds per term) was not significantly longer than that for Read Codes.

Conclusions Clinical Terms Version 3 performed significantly better than Read Codes 5 byte set in capturing the meaning of concepts. These findings suggest that improved coding accuracy in primary care electronic patient records can be achieved with the use of such a clinical terminology.

Introduction

The capture of data in electronic health records is expected to improve clinical effectiveness, governance, and outcomes. However, the data collected must be accurate and consistent.[1] To help satisfy this quality requirement, the use of a standardised clinical terminology (a large knowledge based coding scheme) has been advocated.[2] The NHS developed such a terminology called Clinical Terms Version 3, which was first released in 1994,[3] and is currently completing an evaluation of SNOMED Clinical Terms, which is an enhanced product merging Clinical Terms Version 3 with SNOMED RT (a terminology produced by the College of American Pathologists).[4] Although it shares some features with the earlier Read Codes coding scheme, Clinical Terms offers advantages of unlimited hierarchical depth, multiple relationships, pure concepts (concepts represented by unambiguous preferred terms, semantically correct synonyms with no duplication), and the opportunity to add detail with qualifiers (see bmj.com).[5] Despite high levels of computerisation in UK general practice, surprisingly few developers of electronic patient record systems have adopted Clinical Terms; most still use the Read Codes 5 byte set. Implementing the use of a standard clinical terminology is a key element of the shared electronic record of the NHS information strategy, itself a critical component of the government's commitment to modernise the NHS.[6] The costs of such an implementation are likely to be considerable, yet there is little evidence that using a standard clinical terminology in primary care will accrue benefits.

Comparisons of different clinical coding schemes have mainly been conducted by coding experts looking at the schemes' coverage in relation to existing lists of terms. No study has examined whether a clinical terminology improves the performance of coding electronic patient records by practising doctors in primary care.

The main aim of this crossover study was to determine whether Clinical Terms Version 3 provides greater accuracy and consistency than Read Codes 5 byte set for coding electronic patient records by general practitioners.

Methods

Setting and participants

The study was conducted by 10 general practitioners recruited from practices in urban, suburban, and rural environments in Norfolk. The 10 participants (one woman, nine men) had a median age of 47 (range 40-55) and had been qualified for a median length of 24 years (range 11-33). They self ranked their computer literacy using a simple Likert scale: three reported being "very computer literate" and seven reported a

Editorial by Gardner

School of Information Systems, University of East Anglia, Norwich NR4 7TJ
Philip J B Brown *honorary lecturer in healthcare informatics*

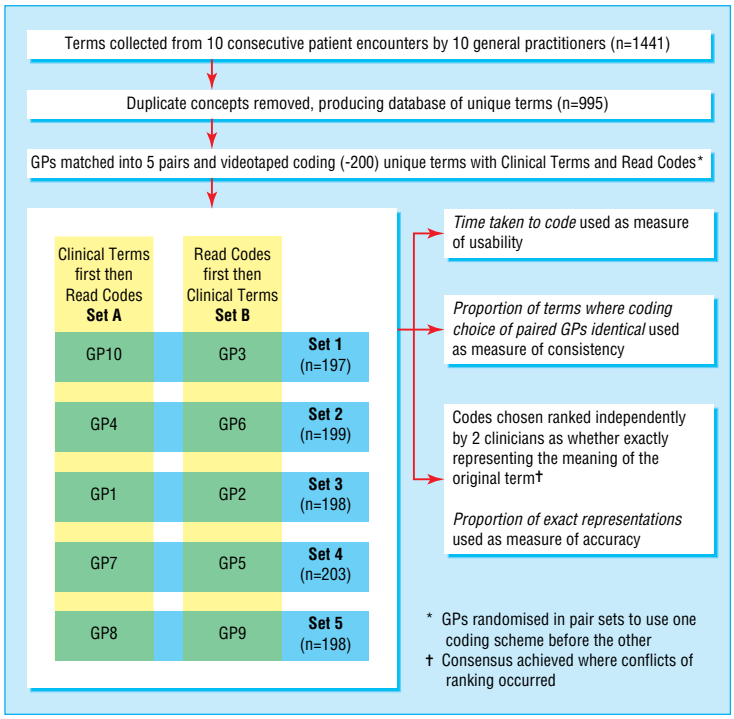
Humbleyard Practice, Hethersett, Norfolk NR9 3AB
Victoria Warmington *research associate*

Bacon Road Medical Centre, Norwich NR2 3QX
Michael Laurence *general practitioner*

Department of Public Health and Primary Care, University of Cambridge, Institute of Public Health, Cambridge CB2 2SR
A Toby Prevost *medical statistician*

Correspondence to: P J B Brown, Humbleyard Practice, Hethersett, Norfolk NR9 3AB
Pjbb@hicomm.demon.co.uk

BMJ 2003;326:1127–30



Summary of study design for comparing coding performance of Clinical Terms Version 3 and Read Codes 5 byte set

"working knowledge." For their frequency of using coding schemes to record consultations, one reported "always," four reported "mostly," and five reported "occasionally." All 10 used the Read Codes 5 byte set; half had not heard of Clinical Terms Version 3, and the rest were aware of it but had no knowledge of its structure and content.

Design

Each general practitioner manually recorded the consultation details of 10 consecutive patients in an arbitrarily chosen consultation session. A simple framework of headings was provided (reason for encounter, diagnosis, treatment, and medical history) to prompt entry of details, but there was no restriction in the terms that could be recorded. The terms from these 100 records were then entered verbatim into an Access (Microsoft) database. We used random number tables to group the general practitioners into five pairs and to randomly select one of each pair to code terms with Read Codes 5 byte set (termed Read Codes in this paper) first and then to code with the Clinical Terms Version 3 (termed Clinical Terms in this paper), and the other doctor in each pair to use the Clinical Terms first followed by Read Codes. We asked the clinicians to code the terms collected from their own records and those from the other doctor in their pair (see figure).

We videotaped each doctor coding his or her allocated file of terms using both Clinical Terms and Read Codes using the same software and hardware. Participants were given standardised instructions and training to identify a code for each term that would be "an acceptable match if the coded record were the only documentation of the concept in a paperless practice."

A researcher reviewed each video and recorded the time taken to code each term. Two researchers then independently examined the coded choices made by each general practitioner and ranked each match as exact or non-exact in representing the meaning of the original term.

Statistical analysis

We used Cohen's κ coefficient to assess agreement among participants in the exactness of coding under each scheme, with a value of ≥0.6 indicating good agreement.[7] We estimated the accuracy of each coding scheme by calculating the proportion of coded choices ranked as being exact semantic representations of the original terms. We estimated consistency by identifying the proportion of cases where the coding choice of the paired general practitioners was the same. We pooled results across general practitioner pairs and calculated confidence intervals using McNemar's test and a bootstrap method (see bmj.com for details).

The length of time taken to find a code was used as a measure of usability of each scheme. We used a non-parametric bootstrap method to estimate 95% confidence intervals and P values for mean time differences.

Results

The 100 consultations generated a total of 1441 terms, providing 995 unique concepts after removal of duplicates. Findings (such as impetigo, upper respiratory tract infection) accounted for 730 of these concepts, with the remaining 265 representing procedures (such as hysterectomy, psychotherapy). The agreement among doctors in the exactness of coding under each scheme was good (see table).

Accuracy of coding schemes

The proportion of concepts ranked as exact semantic representations with Clinical Terms ranged from 0.60 to 0.74 (pooled proportion 0.70) for the 10 participants, with seven of the doctors being in excess of 0.7. By contrast, the proportion of concepts ranked exact with Read Codes ranged from 0.37 to 0.58 (pooled proportion 0.50). All 10 doctors coded significantly more concepts as exact matches with Clinical Terms than with Read Codes ($P<0.001$ for each doctor). The excess proportion of concepts ranked exact with Clinical Terms ranged from 14% (95% confidence interval 7% to 21%) to 27% (19% to 34%) for the 10 participants. The excess proportion of concepts exactly matched with Clinical Terms was similar in the doctors who used this scheme before Read Codes (22%) and in those who used the scheme after using Read Codes (18%), although this relatively small difference represented a significant period effect. We also found a significant carryover effect but the absolute difference was small (see bmj.com)

Performance of 10 paired general practitioners using Clinical Terms Version 3 and Read Code 5 byte set to code terms extracted from 100 consultations. Values are pooled averages over the general practitioner pairs and proportions (95% confidence intervals) unless stated otherwise

Performance measure	Clinical Terms	Read Codes
κ coefficient of interclinician agreement	0.69 (0.64 to 0.74)	0.65 (0.60 to 0.69)
Codes ranked as exact representations of original term	0.70 (0.67 to 0.73)*	0.50 (0.47 to 0.53)
Terms consistently coded	0.58 (0.55 to 0.61)*	0.36 (0.33 to 0.39)
Mean coding time (seconds)	30.2 (28.6 to 31.9)*	36.1 (34.3 to 37.9)

*P<0.001 for difference in performance between schemes.

Consistency of coding schemes
The percentage of concepts ranked consistent (that is, exact matches and coded identically by both members of a pair) ranged from 53% to 63% for Clinical Terms and from 31% to 43% for Read Codes. The excess proportion ranked consistent with Clinical Terms ranged from 21% to 23% and was significant for each of the general practitioner pairs. The pooled proportion of consistent matches by general practitioner pairs was 0.58 for Clinical Terms and 0.36 for Read Codes, with a pooled difference in proportion of 0.22 (0.19 to 0.25) ($P < 0.001$).

Usability of coding schemes
The median coding time for each of the 10 participants ranged from 14 to 27 seconds for Clinical Terms and from 18 to 49 seconds for Read Codes. Compared with Read Codes, the mean excess time taken to code with Clinical Terms ranged from −29 to 12.3 seconds for the pairs of participants. The mean time taken to code with Clinical Terms was shorter by a mean of 5.9 seconds (4.0 to 7.9). However, on the basis of the 850 terms with full data available, there were significant period and carryover effects.

Discussion

Clinical Terms Version 3 performed significantly better than Read Codes 5 byte set in capturing the meaning of concepts required to describe the electronic records of 100 primary care patients. Furthermore, the better accuracy and consistency were accompanied by a shorter average browsing time to find the required coded terms.

Strengths and limitations of our study
We compared the content and usability of the two coding schemes in a practical setting, where clinicians had a variable degree of competency in coding. While formulating the study, we considered videotaping the coding process during live patient consultations. We rejected this in favour of a randomised crossover trial as consistency between experimenters would have been difficult to assess and confounding variables such as time constraints on searching would have been difficult to control. Coding performance and times are therefore only proxy estimates of use in real patient encounters. Further improvement of data entry might be achieved with more sophisticated software than was used in our study—such as by using templates for data entry and menus to access commonly used terms.

We compared Clinical Terms Version 3 with Read Codes 5 byte set rather than the earlier Read Code 4 byte set, which is still in use, because an earlier study had indicated that Read Codes 5 byte set was superior in coverage than the earlier scheme.[8]

The carryover effect in the proportion of terms exactly matched was small compared with the size of the difference between the two coding schemes in each period. The carryover effect for coding time reflected the change from the first period to the second period in the difference between the schemes. Four of the five doctors who coded first with Read Codes took more than 10 seconds longer to code each term than they did with Clinical Terms; review of the video comments of the remaining participant suggested that the longer coding time in the second period related to user fatigue (including remarks about the doctor's uncertainty of meaning of the original term and technical difficulties in using the notebook keypad). Only one of the participants who coded first with Clinical Terms took more than 10 seconds longer to code with this than with Read Codes, and this may be accounted for by the doctor's familiarity with the content of Read Codes. The small number of participants limits our ability to explain such differences with certainty, and we have cautiously interpreted them to say that the time taken to code with Clinical Terms was not significantly longer than that with Read Codes (see bmj.com).

We did not try to measure the potential clinical importance of the non-exact matches. Judging the importance of non-exact matching would have introduced a further subjective element, requiring further checks of inter-rater reliability that were outside the scope of the study, although our data provide valuable material for further study.

What is already known on this topic

Clinical terminologies such as Clinical Terms Version 3 have been shown to offer greater coverage than earlier coding schemes when used by experts for coding data in electronic patient records, mainly by virtue of their larger size

Implementing the use of a terminology in a health service will be costly, and evidence is needed that this will improve the quality of medical data

What this study adds

When used by general practitioners Clinical Terms Version 3 performed significantly better than Read Codes 5 byte set in consistently coding the meaning of concepts for electronic patient records in primary care

Despite the larger size of Clinical Terms, the improvement in accuracy was achievable without an increase in the average browsing time to find the required coded terms

Improved coding accuracy and consistency in primary care electronic patient records can be achieved with Clinical Terms Version 3

Conclusion
The coding of clinical records is an important aspect of medical audit, research, epidemiology, management of resources, and the direct care of patients. For information technology to be fully adopted, clinical notions that are often complex must be accurately and easily represented as coded concepts that are "user friendly" and easily retrievable. Our study suggests that substantial advantages may be achieved by investing in the implementation of Clinical Terms Version 3 or a similar terminology.

We thank Ian Harvey (Health Policy and Practice, University of East Anglia) for his advice in designing the study and Doreen Cochrane (SuNet facilitator, Health Policy and Practice, University of East Anglia) for invaluable input during the preparation of this proposal. The following clinicians participated in the study: Robert Bawden, Peter Burrows, Jamie Dalrymple,

Stephen Daykin, Christopher Hand, Carly Hughes, Andrew Leaman, David Munson, Tony Press, and Rob Stone.

Contributors: see bmj.com

Funding: The study was funded by the NHS Eastern Region R&D grant No RCC33031.

Competing interests: PJBB advises the NHS Information Authority on coding and terminology.

Ethical approval: The study was granted ethical approval by the Norfolk and Norwich Ethical Committee.

1 Brown PJB, Sönksen P. Evaluation of the quality of information retrieval of clinical findings from a computerised patient database using a semantic terminological model. *J Am Med Inform Assoc* 2000;7:401-12.
2 Evans DA, Cimino JJ, Hersh WR, Huff SM, Bell DS, for the CANON Group. Towards a medical-concept representation language. *J Am Med Inform Assoc* 1994;1:207-17.
3 Severs MP. The clinical terms project. *Bull R Coll Physicians Lond* 1993;27(2):9-10.
4 Stearns MQ, Price C, Spackman KA, 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.
5 O'Neil M, Payne C, Read JD. Read codes version 3—a user led terminology. *Methods Inf Med* 1995;34:187-92.
6 NHS Executive. *Information for health: an information strategy for the modern NHS.* London, Department of Health, 1998.
7 Altman DG. *Practical statistics for medical research.* London: Chapman and Hall, 1996.
8 Brown PJB, Odusanya L. Does size matter?—Evaluation of value added content of two decades of successive coding schemes in secondary care. In: Bakken S, ed. *Proceedings of the 2001 AMIA Fall Symposium.* Philadelphia: Hanley and Belfus, 2001:71-5.

(Accepted 5 March 2003)

ELPS

This is an abridged version; the full version is on bmj.com

Use of Read codes in diabetes management in a south London primary care group: implications for establishing disease registers

Jeremy Gray, Douglas Orr, Azeem Majeed

Editorial by Gardner

Battersea Research Group, Bolingbroke Hospital, London SW11 6HN

Jeremy Gray *director*

Douglas Orr *researcher*

Primary Care Research Unit, School of Public Policy, University College London, London WC1H 9QU

Azeem Majeed *professor of primary care*

Correspondence to: J Gray

BMJ 2003;326:1130–2

Abstract

Objective To establish current practice in the use of Read codes for diabetes.
Design Cross sectional study.
Setting 17 practices in the Battersea primary care group in southwest London.
Data sources Computerised medical records.
Main outcome measures Number of codes in use in all practices; variation in the use of codes between practices; and prevalence of Read code use in diabetic patients.
Results At least 9 separate Read code groupings and 25 individual diabetes codes were in use in the 17 general practices. Only one Read code (C10, diabetes mellitus) and its subcodes was being used in all 17 practices, but its use varied from 14% to 98% of patients with diabetes. The use of other key Read codes for monitoring the care of patients with diabetes also varied widely between practices; for example, $<20\%$ of practices used the code for the location of care. Less than half of patients (45%) with diabetes had their type of diabetes coded, and even fewer (21%) had measures such as the examination of the retina coded.
Conclusions The use of Read codes for diabetes needs to be standardised and coding levels improved if valid diabetic registers are to be constructed and the quality of care is to be monitored effectively. Until all patients with diabetes have the C10 Read code recorded, clinicians will have to use a wide range of Read codes and prescribing data to ensure that diabetes registers are complete.

Introduction

Diabetes is associated with considerable morbidity and mortality, and improving its management is a priority in England and Wales.[1 2] The government envisages a much greater role for primary care in managing diabetes[3 4] and has launched several initiatives. These include an expansion of services for patients with diabetes through greater investment in primary care, the creation of specialist general practitioners,[5] baseline assessments of current services in primary care trusts to help plan services,[6] a draft national specification for diabetes registers,[7] and improvement in patient care through investment in information and communication technology.[8]

Accurate disease registers will be essential to improve the care of people with diabetes.[9] Read codes were developed in the 1980s and are currently used to code clinical data in primary care in the United Kingdom.[10] New codes are released regularly by the NHS Information Authority. In addition, producers of general practice clinical computer systems can add their own codes, as can individual practices.

The box gives an example of a Read code hierarchy. The Read coding system is complex, and a disease can be coded in many different ways—for example, through a specific disease code, history and symptoms, or investigations and procedures. This can lead to wide variations in the way in which general practitioners code clinical problems. We examined the Read codes used in recording information on the management of diabetes in one primary care group, as the first step in developing a local disease register.

Methods

We used a two stage process to identify Read codes currently being used to record the management of diabetes in 17 general practices in the Battersea Primary Care Group in southwest London. Firstly, we tried to identify all patients with diabetes and all the Read codes associated with their management. We then calculated the proportion of patients for whom each code was used. All 17 practices used the EMIS computer system.