Mnemonic coding system for clinical data entry into laboratory computers: Its effect on quality and efficiency

I R Bailey, K B Page, R G Jones, R B Payne, A J Little

Abstract

A simplified coding method for entering the clinical details found on pathology request cards was developed. The method uses a basic four letter code, derived from the initial character of the first four words in a clinical detail, being expanded to four characters with letters from the final word if the number of words is less than four. Rules were devised to cope with common medical terminology. In excess of 90% of clinical details on request cards are readily input by clerical staff using our coding system, and 8% of clinical details are used intelligently by the computer in scheduling further tests or automatically commenting on results.

A carefully designed coding system such as the one outlined above could greatly facilitate input of clinical detail without the penalty of reduced throughput.

The recent acquisition by the NHS of the Read system of clinical coding¹ is a welcome move towards standardisation of computer coding systems for medical records.¹⁻³ It promises to facilitate long term storage, recall and communication of medical data between health care units thus improving the management and use of scarce health care resources.⁴

This is not the only consideration for many hospital service departments, where the numbers of patients may be very large and a limited amount of clinical data are required for a short period. A particular example is the clinical chemistry laboratory in which clinical details are required for the appropriate handling of samples in the same period immediately before, during, and after analysis.

Request cards arrive in the laboratory with the patient's demographic data, the tests required, and a highly heterogeneous mix of clinical details, varying in length, legibility, and degree of sophistication. The value of clinical details is that although they may not accurately indicate the final diagnosis, they usually indicate the reason for the test or the question the clinician is asking. They may suggest that additional laboratory tests need to be performed to establish the patient's true biochemical state, particularly in the case of endocrine investigations, or indicate the urgency of a request when the usual local arrangements for urgent tests have broken down in some way. The validation of results

produced by the laboratory also requires clinical data for the selection of results to be telephoned, the addition of further tests as indicated, and for the detection of one-off errors ("fliers") and other blunders, such as specimens taken from the wrong patient.

One of the problems encountered in the computerisation of a laboratory is how to retain the clinical data.⁵ If it is not entered into the computer, additional test scheduling, as used by many laboratories for thyroid function tests—for example, cannot be implemented without referring back to the original request form, and automatic interpretive comments cannot be based on relevant clinical information. Frequent recourse to request cards must also be made during the clinical validation procedure. If the clinical details are entered as free text the entry of requests is slowed and may delay the reporting of results. Free text entry leads to spelling mistakes and benefits only manual validation procedures, automatic procedures requiring coded information.

If a coding system is used it has to be able to cope with questions, medical and vernacular terminology, be comprehensive and easy to use.¹⁴ Although menu driven data entry facilities are commonly provided on modern medical computing systems—for example, for use with numerical and alpha-numeric coding systems—they are never as quick and easy to use as a system in which the codes are readily predictable from the information on the request form. Such a system allows clinical details to be entered with a minimum of time and effort and allows a high throughput to be maintained, without knowledge of disease associations. Predictability depends on the design of the system, and careful design should result in a minimum number of easily remembered rules.

Methods

The coding system was developed on a clinical laboratory data management system supplied by Telepath Systems Ltd (Birmingham, England) running on a PDP 11/83 with an ethernet link to a PDP 11/73 computer. Clinical details may be entered either as a code or as free text. Menu driven search facilities can be used to find appropriate codes. The coding system went through two prototype stages. Our initial coding system was based on using the shortest code possible, but this meant that codes were of variable length (one to five letters) and it proved difficult to predict the

Department of
Chemical Pathology,
St James's University
Hospital, Beckett
Street, Leeds LS9 7TF
I R Bailey
K B Page
R G Jones
R B Payne
A J Little

Correspondence to:
Dr I R Bailey, Department
of Chemical Pathology,
Farnborough Hospital,
Farnborough, Kent
BR6 8HD.

Accepted for publication
12 June 1991

codes. On review of the codes in the original list we settled on a four letter code as the basic code length for our second version, although Telepath Systems (TPS) allows a maximum clinical detail code length of six characters.

For each clinical detail there is a code and an expansion (up to 35 characters) which is usually identical with the clinical detail as received on the card. This is important because clerical officers can then easily recognise the appropriate code by its expansion when the latter appears on the computer screen, thus reducing errors and increasing confidence. The exceptions are those cases where there are insufficient characters in the expansion field to display the clinical detail in full, and in these cases abbreviations may be used. All clinical details appear on worksheets and are returned on laboratory reports as the expansion.

RULES

The first and basic rule is that a code is derived from the initial letters of the words in a clinical detail phrase, the code being made up to four letters by taking subsequent letters from the final word if required—for example, superior vena caval obstruction is coded SVCO and acute abdomen, AABD. It follows, therefore, that in the case of a one word clinical detail that its code is the first four letters of the word—for example, acromegaly is coded ACRO. Only if the final word is too short to generate a four letter code does a shorter code result—for example, hepatitis A is coded HA. Hyphenated words are treated as two words. The words “and”, “in”, “than”, “of” and “of the” are ignored for the purposes of coding, but “on” is not, except when the dose of a drug is given. Thus carcinoma of the ovary is coded as COVA and its expansion is carcinoma of the ovary, but “on digoxin” is coded ODIG.

The second rule is that an abbreviation used in a clinical detail is treated as one word for the purpose of deriving the code and also appears in the expansion as the abbreviation—for example, “CRF, on dialysis,” is coded CODI. Where abbreviations are less than four characters the code is the length of the abbreviations—for example, DM is coded DM. For very common abbreviations longer than four characters the whole abbreviation is used—for example, NIDDM.

The third rule is that numbers are used in their entirety and that the code length is extended to include both the number and one further character. Units of dose are ignored, but dose frequency is not. Thus “on lithium 1000 mg bd” is coded L1000B, the maximum six character code length.

The fourth rule states that queries are indicated by a Q in the code. Thus the code for “? myeloma” is QMYEL, and that for “anaemia? cause” is ANAEQ. Codes containing a Q as a suffix or prefix therefore consist of five letters. Details such as “renal transplant? rejecting”, however, are coded like RTQR.

Certain prefixes and suffixes are common in medical terminology. From the clinical details and coding experience of our initial version the prefixes likely to cause problems and the common suffixes were identified. The fifth rule then followed. This rule uses abbreviations for common prefixes such as hyper- and hypo-, suffixes such as -ectomy and -ostomy, and common final words such as disease. A list of the prefixes is given in table 1 and common final words in table 2. The suffix -ectomy is shortened to X and -ostomy to Y. When a word has both a prefix and a suffix, only the suffix is used, because otherwise short codes would result. Thus hepatectomy is coded HEPX and not HX. If a single word clinical detail occurs that would normally be abbreviated, either the first four letters or its abbreviated code can be used—for example, ANAE or A for anaemia, PAIN or P for pain.

Where a long hand clinical detail coded by these rules would have a code identical with its common abbreviation—for example, CABG for coronary artery bypass graft—the abbreviation takes precedence and becomes the code. The long hand clinical detail has the code extended by one letter, the next letter of the final word, in this case becoming CABGR. All common medical abbreviations of four letters are treated in this way. Where a unique code cannot be derived from the above rules the code is extended to five letters by the next letter in the clinical detail that makes it unique. APAT is the code for acute pancreatitis, so the code for alcoholic pancreatitis becomes APATI. Although this last rule reflects a failure of the system to be entirely predictable, in our experience these codes account for only 2.5% of the total. The commonest occurring detail is always given the four letter code.

The rules are sufficiently rigid for us to have been able to develop a program on a personal computer which automatically generates a unique code from the free text. This enables us to convert free text clinical detail entries to

Table 1 Abbreviations used for common prefixes to construct codes

Prefix	Abbreviation	Example of clinical detail	Code
arthr	A	Arthropathy	AOPA
bronch	B	Bronchitis	BITI
cysti	C	Cystinosis	CNOS
dys	D	Dyspnoea	DPNO
haem	H	Haematemesis	HATE
hepat	H	Hepatomegaly	HOME
hyper	HE	Hypertension	HETE
hypo	HO	Hyponatraemia	HONA
neph	NEP	Nephropathy	NEPO
neuro	N	Neuroblastoma	NBLA
osteo	O	Osteomalacia	OMAL
pancre	P	Pancreatitis	PATI
poly	P	Polydipsia	PDIP
post	PO	Post-op	POOP
pre	PE	Pre-dialysis	PEDI
previous	PR	Previous surgery	PRSU
xanth	X	Xanthelasma	XELA

Table 2 Abbreviations used for common final words

Word	Abbreviation	Example of clinical detail	Code
Abscess	A	Perianal abscess	PERA
Anaemia	A	Microcytic anaemia	MICA
Carcinoma	C	Breast carcinoma	BREC
Coma	C	Diabetic coma	DIAC
Disease	D	Cushing's disease	CUSD
Failure	F	Adrenal failure	ADRF
Infusion	I	Oxytocin infusion	OKYI
Overdose	O	Paracetamol overdose	PARO
Pain	P	Bone pain	BONP
Stenosis	S	Mitral stenosis	MITS
Stopped	S	Diuretics stopped	DIUS
Syndrome	S	Cushing's syndrome	CUSS
Tumour	T	Bladder tumour	BLAT

codes which can be added to the dictionary.

Combined details on request cards are also a problem as the number of potential combinations is large. As the TPS system is currently restricted to a single expansion a further rule which stated the order of priority for data entry was necessary. The order of priority chosen was as follows: clinical details referring to treatment—for example, “on phenytoin”, had the highest priority—followed by details referring to thyroid disease, then those referring to lipid disorders, renal disease and lastly diabetes; these details were more likely to be used by the system. Only one detail was entered as a code, based on the order of priority, remaining details being entered as free text in another field usually used for specimen comments. Common combinations, however, were coded—for example, “CRF, diabetic” is coded CDIA. Abbreviated details are taken first so that Diabetic, CRF is also coded CDIA. A future upgrade to the TPS software is expected to allow for automatic combination of the expansions of multiple codes.

To support the use of generic rather than proprietary names for drugs, when a request card has a proprietary name such as “on Zocor” we have coded this OZOC, but the expansion is “on simvastatin (Zocor)”; “on simvastatin” is coded OSIM and the expansion is “on simvastatin”.

DATA COLLECTION

The numbers of clinical details coded were obtained by reviewing all clinical detail data entered into the laboratory computer database over periods of three or four months. The number of code entries requiring the use of search facilities was obtained by observing users of the system under normal data entry conditions.

Results

In the first four months following the introduction of the TPS system with this coding system for clinical details, 64 832 requests were received in the laboratory. Of these, 88% were entered by our clerical staff using coded clinical details. At the end of this period about 71% of all cards were being coded without recourse to search facilities. A further 5% of cards had no clinical details given or were illegible, this information being entered as the codes NCDG or ILLE instead. Table 3 shows the effect of time on the number of clinical details coded. The number of cards without clinical details or illegible details remains at 5%. We have found that a directory of 2500 codes covers most clinical details seen on our request cards. The most common codes used are shown in table 4. We have found that when searching is required the system is very efficient. A search through the codes starting with the first three letters of the expected code, on theoretical grounds, will find the desired code within the next 26; in practice this is usually within the first five.

Of clinical details entered, 8.4% are used intelligently by the computer. They are used to ensure that automatically appended inter-

Table 3 Request cards coded by our system

	Coded	Requiring searches
First four months	88%	29%
Second four months	95%	10%
Most recent three months	98%	ONMA

ONMA = observations not made.

pretative comments are appropriate to the clinical context—for example a high cholesterol result would normally have a comment added depending on the magnitude of the abnormality, but if the clinical detail was, for instance on bezafibrate no comment would be appended. They are used to trigger further tests—for example, the addition of a TSH assay to a T4 request if the clinical detail is on thyroxine.⁶ They also trigger the computer to use clinically appropriate limits for listing the results on an abnormal results list for checking by the duty biochemist or chemical pathologist—for example, to detect an unexpectedly low serum creatinine concentration if the clinical detail is chronic renal failure.

Discussion

Our experience shows that a coding system for clinical details which relies on easily predictable codes can facilitate the entry of clinical details into the laboratory computer without appreciably increasing the work of the clerical officers. Assuming that the retention of clinical details is a required function, we estimate that the use of our coding system saves the equivalent of two hours data entry time per day when compared with a free text entry system. There are additional savings which are difficult to quantify through the availability of clinical data at all work stations. The clinical data can also be easily searched for retrospective review. As the data are entered more rapidly, the work of the laboratory is completed earlier and reports returned to clinicians faster. Earlier experience in our own laboratory and that in other local laboratories who have recently installed the TPS or a similar system highlights the problems caused when a heterogeneous coding system is allowed to develop. In laboratories where several people have been involved in devising codes using independent rules the search facilities are used far more often than is necessary and many details are not entered, with a consequent reduction in efficiency.

At present over 95% of clinical details on request cards can be coded by our system, about 90% without recourse to search facilities despite the fact that it is neither entirely comprehensive nor are the codes entirely

Table 4 Most frequently used codes

Code	Expansion	Number	Percent
RTRA	Renal transplant	3037	4.7
QMI	? MI	1641	2.5
TRAN	Transplant	1429	2.2
PEOP	Pre-op	1358	2.1
POOP	Post-op	1257	1.9
DM	DM	1139	1.7
DIAB	Diabetes	1017	1.6
ODIU	On diuretics	955	1.5
CPAI	Chest pain	824	1.3
ANAE	Anaemia	712	1.1

predictable. We believe, however, that a carefully designed system such as ours can greatly facilitate data entry without the penalty of reduced throughput. We are currently considering the development of a background computer translation which could convert a code entered by our system to its equivalent Read code to permit standardisation of reporting for case mix management and audit purposes.

- 1 Chisholm J. The Read clinical classification. *Br Med J* 1990;300:1092.
- 2 McDonald CJ, Hammond WE. Standard formats for electronic transfer of clinical data. *Ann Intern Med* 1989; 110:333-5.
- 3 Earlam R, Körner, nomenclature and SNOMED. *Br Med J* 1988;296:903-5.
- 4 Read J, Benson T. Comprehensive coding. *Br J Healthcare Comp* 1986;3:22-5.
- 5 Hercz L, Laszlo ChA, Reesal MR. A computerised information system for pathology. *Meth Inform Med* 1975;14: 181-8.
- 6 Little AJ. Automation of a laboratory thyroid function tests strategy. *TeleForum* 1990;1:4-5.