Ann R Coll Surg Engl 1998; **80**: 293–296

A novel method for the assessment of the accuracy of diagnostic codes in general surgery

N Gibson PhD
Research Assistant

S A Bridgman MD FRCS
Senior Lecturer in Surgical Trials and Epidemiology

School of Postgraduate Medicine, University of Keele

Key words: Appendicitis; Surgery; Operative; Information systems; Data collection/standards; Medical records system, computerised

The aim of this study was to describe the accuracy of diagnostic coding in general surgery in a district general hospital, the North Staffordshire Hospital NHS Trust (NSHT), Stoke-on-Trent.

An assessment was carried out by comparison between codes ascribed by hospital coders and expert external coders.

Patients who had a finished consultant episode (FCE) in the specialty of general surgery at NSHT were included in the study.

The sampling frame was general surgery FCEs at NSHT purchased by North Staffordshire Health Authority (NSHA) with an episode end date between 1 May 1995 and 31 December 1995. Every 15th record was sampled. Of 455 records sampled, 157 (35%) were in active use and were excluded but not replaced; therefore, 298 (65%) records were studied in detail.

Outcome was measured by the accuracy of primary diagnostic codes ranked 1, 2, 3, 4, from highest to lowest levels of inaccuracy; a description of where errors occurred in the data cycle was recorded.

Errors were found in 87/298 (29%) records; 25/298 (8%) records had an error at the highest level (ie wrong ICD-10 chapter), and 44/298 (15%) at the third level. Of the errors, 68/87 (78%) occurred between the medical record and the admission form. A substantial percentage (29%) of records had inaccurate diagnostic codes.

It is concluded that coding should be carried out from the medical record rather than from the admission form (KMR1). The proportion of records with errors suggests that a routine data coding audit would be useful to improve the accuracy of routine diagnostic codes.

Correspondence to: Mr S A Bridgman, Springfields, Jervis Lane, Meaford, Stone, Staffordshire ST15 0PZ

Parsaye and Chignell *(1)* observe that organisations seem to wait until disaster strikes before discovering the importance of data quality. There is good evidence that the accuracy of clinical data in health information systems in the UK is a major problem *(2,3)*. Drennan *(4)* found that 32% of primary diagnostic codes and 17% of primary procedures codes were incorrect out of a total of 2404 records surveyed in four acute care hospitals for the financial year of 1990/1991.

The traditional method of measuring accuracy determines whether a data item is either correct or incorrect. However, International Classification of Diseases (ICD) *(5)*, revision 10 codes are hierarchical in nature, with the higher the level of error the more fundamental its effect. Therefore, in addition we measured data accuracy according to the level of inaccuracy of the error. As measurement of the accuracy of existing data alone does not improve subsequent data quality *(6)*, we also observed where in the data life cycle errors occurred, with the aim of suggesting how data quality could be improved.

Method

Sample

North Staffordshire Health Authority has 483 000 residents and has an acute contract with one predominant provider. The sample of inpatient records for the study was taken from North Staffordshire Health Authority's (NSHA) contract management system. Each record in this database relates to a finished consultant episode (FCE) which is defined as "the time a patient spends in the continuous care of one consultant using a hospital on the site of one health care provider" *(7)*. There was no previous knowledge of the level of accuracy in the codes at the North Staffordshire Hospital NHS Trust (NSHT). From the limited information available, a sample of approximately 455 database records was judged a reasonable and practical size to study given the resources available. The size of the sample took into account that a proportion of this sample would not be used if the patient's medical record was in active use.

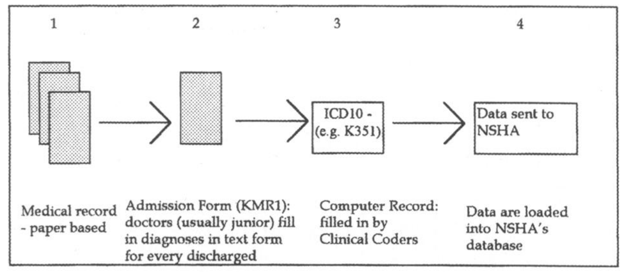
The sampling frame was FCEs in the specialty of general surgery at NSHT with an episode end date between 1 May 1995 and 31 December 1995 and with NSHA as the purchaser. General surgery was chosen because, as a specialty, it has a wide variety of diagnoses and codes. The database records are held in the order they were entered. Every 15th database record was sampled. Medical records staff at NSHT then retrieved the selected medical records. Records in active use at the time of the retrieval were excluded and not replaced.

Data life cycle at North Staffordshire Hospital

To understand this study, it is helpful to consider the data life cycle at North Staffordshire Hospital. The hospital has a manual system of medical records for day-to-day management of patients. Upon discharge of a patient a doctor (usually a junior) writes the diagnoses and procedures a patient has undergone in hospital on an admission form, the KMR1 (Korner Monitoring Return 1). This form is sent to the hospital clinical coders whose job it is to assign ICD-10 for each patient. These data, along with other elements of the standard National Health Service (NHS) contract minimum dataset, are then sent to NSHA. This cycle is referred to as the data life cycle (Fig. 1).

Classification of level of error

Comparison was made between the NSHA database record and the medical record by two expert coders (from Trent Clinicode NHS, Doncaster Royal Infirmary) as the gold standard. Records were considered to have an



ICD = International Classification of Diseases; NSHA = North Staffordshire Health Authority; KMR1 = Korner Monitoring Return 1

Figure 1. Data life cycle at North Staffordshire Hospital Trust (NSHT).

error or no error, but in addition those with an error were ranked from 1 (highest level of error) to 4 (lowest level of error). This method of measurement can be explained by examining the structure of ICD-10 codes, which are grouped into categories which form a hierarchy of coding levels (Table I).

For example, if the database record code was K389 but it should have been K350, then the error would be classified as being at the third level. A code inaccurate at the chapter of related codes, or first level (eg J350 instead of K350) can be considered to have a higher level of inaccuracy than a code which is only in error at the third level (eg K389 instead of K359).

Where in data life cycle errors occurred

The expert coders studied the admission forms (KMR1) and medical records separately and ascribed diagnostic codes to the records based on information from both these sources. The code ascribed from expert coders from the medical records was taken as the gold standard. Comparisons were then made between:

- code on NSHA database record with code ascribed by expert coder from admission form (KMR1);
- code ascribed by expert coder from admission form with code ascribed by expert coder from medical record notes.

The proportion of errors were also described by admission type, ie whether the patient was an elective day case, elective inpatient, non-elective admission who underwent an operation or procedure, and non-elective admission who did not undergo an operation or procedure.

Table I. Hierarchy of ICD-10 coding levels with example for each coding level

Hierarchy of coding levels	*Example code*	*Description*	*Error classification*
Chapter of related codes	K00–K93	Diseases of the digestive system	First level (1)
Group of three-character codes	K35–K38	Diseases of the appendix	Second level (2)
Three-character code	K35	Acute appendicitis	Third level (3)
Four-character code	K350	Acute appendicitis with generalised peritonitis	Fourth level (4)

Table II. Admission type and hierarchical coding level of error

Admission type	*Hierarchical coding level of error* 1st level	2nd level	3rd level	4th level	Total errors	% Incorrect primary diagnosis errors	95% Confidence interval
Elective inpatient	7	1	13	3	24/83	29%	19% to 39%
Elective day case	9	1	8	6	24/99	24%	16% to 33%
Non-elective—procedure	3	2	15	3	23/53	43%	30% to 57%
Non-elective—no procedure	6	0	8	2	16/63	25%	14% to 36%
Total	25 (8%)	4 (1%)	44 (15%)	14 (5%)	87/298	29%	28% to 31%

Results

Of the sample of 455 database records, 157 medical records were in active use and therefore excluded from the study. Thus, 298 records were studied in detail. Of these, 83 (28%) cases were elective inpatients, 99 (33%) elective day cases, 53 (18%) non-elective cases who underwent a procedure and 63 (21%) were non-elective cases who did not undergo a procedure.

Incorrect primary diagnostic codes were observed in 29% (87/298) of cases. A higher proportion of errors; tended to occur in non-elective cases that underwent a procedure, with 14/23 of these errors from the coding of acute appendicitis (K35) as unspecified appendicitis (K37) (Table I and Table II). In all, 25 errors (8% of sample) occurred at the highest level of the hierarchy, ie they were coded from the wrong chapter of diseases. About one-half of the errors occurred at the third hierarchical level (Table II).

There was an error between transcription of information between the medical record and admission form in 23% (68/298) of records and between the admission form and database record in 7% (22/298) of records. There were three cases where errors occurred at both stages in the data life cycle (Table III).

Table III. Frequency of errors at each stage in the data life cycle

Patient classification	*Error between medical record and admission form*	*Error between admission form and database record*	*Sample size*
Elective inpatient	22 (27%)	2 (2%)	83
Elective day case	15 (15%)	9 (9%)	99
Non-elective—procedure	21 (40%)	4 (8%)	53
Non-elective—no procedure	10 (16%)	7 (11%)	63
Total	68 (23%)	22 (7%)	298

Validity of results

Theoretically, the exclusion of 35% of cases in active use may have altered the case mix. However, we still believe that the conclusions on data quality from this study are generalisable, as the exclusion of cases was not on the basis of their data quality. It is assumed that the medical records and the diagnoses written in the medical records are accurate. It is also assumed that the external coders who reassigned the codes were accurate and not in error. Another assumption is that the information in the medical records available to the doctors who filled in the admission form, and to the expert clinical coders was the same but, as discussed below, this may not necessarily be the case.

Discussion

This study has shown that nearly one-quarter of FCEs in general surgery in NSHT had inaccurate primary diagnostic codes ascribed, and that 8% of FCEs were given codes in different ICD-10 chapters than that in which they should have been (Table II). The majority of errors occurred between the medical record and admission form (Table III). While this work confirms other work showing that hospital statistics have a high level of inaccuracy, it goes further than previous work we are aware of in showing that just over one-quarter of the errors had a high level of inaccuracy, so that patients were not classified in the appropriate chapter of ICD-10 codes. Such inaccuracy is important and this study casts doubt on the validity of routine statistics derived from hospital general surgery activity.

Significance of level of coding

Data accuracy is a 'measure of agreement with an identified source' *(8)*. This methodology used in the study was novel, as previous studies have simply counted records as correctly coded or incorrectly coded *(2,3)*. Measurement of the level of error is useful as data have a myriad of uses and different levels of error may have a different significance depending on the use. For example, for a surgical unit reviewing their cases of acute appendicitis (K35), they will miss cases if coding is inaccurate at the third level. However, an audit of all cases involving the appendix will enable identification of cases even if the classification of errors is at the third level.

Reduction of proportion of records with errors

The majority of errors occurred between the medical record and the admission form, which is written by junior doctors on discharge of a patient (Table III). There were fewer errors between the admission form and the database record (ie coding errors). This would suggest that coding directly from medical records should improve data accuracy. In particular, the medical records include the discharge letter in which a more senior doctor often states the final diagnosis, and in this study the expert coders did find these letters very helpful. However, if the reason for inaccuracy was that pathology reports were not available, then accuracy may not improve unless these reports were available to the coder. This effect of the availability of pathology reports on data accuracy requires further study.

Need for quality control in coding

The substantial number of records in which coding was inaccurate suggests the need for the introduction of a systematic data quality audit system in NSHT, at least in general surgery. The problem of inaccurate data is not new, and the inaccuracy found in this study is very similar to that found by Drennan *(4)*. However, if the accuracy of diagnostic coding in NSHT is reflected in other hospitals and other specialties, it would cast doubt on the validity of trend and other surveillance data used in the National Health Service. One of us still remembers working as a junior general surgeon when the firm received a report detailing the 30 caesarean sections it had performed. If routine health service statistics cannot be trusted, then the public and health professionals alike will have good reason to be sceptical of decisions made on the basis of them.

Further research

The expert clinical coders coded from information available in the medical records. An extension to this study would be useful if another validation arm included the surgeon in charge of the case coding records with the expert coders, and comparing codes with those obtained in the initial study. It would also be useful to extend the study to other hospital specialties and other hospitals.

We acknowledge the work of Sue Douglas and Denise McCafferty from Trent Clinicode NHS, Doncaster Royal Infirmary and the advice Mr M Deakin, Senior Lecturer in Surgery, and Professor J Elder, Professor of Surgery, at the University of Keele, Dr Pearl Brereton, Department of Computer Science, University of Keele. We are also grateful to Medical Records staff at NSHT for retrieving the medical records.

References

1 Parsaye K, Chignell M. Quality unbound. *Database Programming and Design* 1995; **8**: 38–42.

2 Mukherjee AK, Leck I, Langley FA, Ashcroft C. The completeness and accuracy of Health Authority and Cancer Registry records according to a study of ovarian neoplasms. *Soc Public Health* 1991; **105**: 69–78.

3 Barrie JL, Marsh DR. Quality of data in the Manchester orthopaedic database. *BMJ* 1992; **304**: 159–62.

4 Drennan Y. Data quality, patient classification systems and audit: a recent study. In: Richards B *et al.*, eds. *Current Perspectives in Healthcare Computing*. Weybridge: BJHC Books, 1994.

5 World Health Organisation. *International Statistical Classification of Diseases*, 10th Revision. Geneva: WHO, 1992.

6 Huh YU, Keller FR, Redman TC, Watkins AR. Data quality. *Information and Software Technology* 1990; **32**: 559–65.

7 NHS Executive Information Management Group. *NHS Data Dictionary*. Version 2.0. Birmingham: NHS Executive, March 1995.

8 Detrekoi A. Data quality management in GIS systems. *Comput Environ Urban Systems* 1994; **18**: 81–5.

Received 27 November 1997