Comparative hospital databases: value for management and quality

Robert Cleary, Richard Beard, James Coles, Brendan Devlin, Anthony Hopkins, Dale Schumacher, Iden Wickings

Abstract

Objectives—To establish an accurate and reliable comparative database of discharge abstracts and to appraise its value for assessments of quality of care.

Design—Retrospective review of case notes by trained research abstractors and comparison with matched information as routinely collected by the hospitals' own information systems.

Setting—Three district general hospitals and two major London teaching hospitals.

Patients—The database included 3905 medical and surgical cases and 2082 obstetric cases from 1990 and 1991.

Main measures—Accessibility of case notes; measures of reliability between reviewers and of validity of case note content; application of high level quality indicators.

Results—The existing hospital systems extracted insufficient detail from case notes to conduct clinical comparative analyses for medical and surgical cases. The research abstractors at least doubled the diagnostic codes extracted. Inter-abstractor agreement of about 70% was obtained for primary diagnosis and assignment to diagnosis related group. These data were sufficient to create a comparative database and apply high level quality indicators designed to flag topics for further study. For obstetric-specific indicators the rates were comparable for abstractors and the hospital information systems, which in each case was a departmentally based system (SMMIS) producing more detailed and accessible data.

Conclusions—Current methods of extracting and coding diagnostic and procedural data from case notes in this sample of hospitals is unsatisfactory: notes were difficult to access and recording is unacceptably incomplete.

Implications—Improvements as piloted in this project, are readily available should the NHS, hospital managers, and clinicians see the value of these data in their clinical and managerial activities.

(*Quality in Health Care* 1994;3:3-10)

Introduction

The NHS requires that a range of information, the "provider minimum data set,"¹ is collected on every episode of inpatient care. The data

set, a development of the previous Körner returns, includes basic demographic information; fields relating to referral, admission, and discharge, one diagnosis code or more (according to the International Classification of Diseases); and details of any operative procedures. Generally, the data are gathered by coding clerks reviewing either case notes or, more likely, summaries thereof. Considerable evidence is available concerning the completeness and accuracy of data obtained from hospital case notes in the United Kingdom (UK) through routine abstraction and coding systems.²⁻⁴ Even with the increased emphasis on quality of coding by the requirements of the new contracting arrangements, it remains clear that poor quality data are still inhibiting use of the data set, both for general and clinical management.

Similar problems with quality of data were identified in the United States (US) as early as 1977,⁵ and these were emphasised by the introduction of the prospective payment system.⁶ These have been addressed, and to a large degree solved, in the US by various training programmes, database development activities, software packages, and encoding incentives associated with review of utilisation and hospital payment. As the quality of routine data in the US has increased, so too has interest in using such data to address questions of appropriateness, effectiveness, and the quality of care.⁷ A seminal instance of this approach is the demonstration by Wennberg *et al* of the use of medical insurance databases in comparing the outcomes of prostatectomy associated with different providers and surgical techniques.⁸ Database analysis also has an important place among the battery of techniques being used by the federally funded Patient Outcome Research Teams (PORTs).⁹ These teams are supplementing meta-analyses and prospective data collection with analysis of claims data in assessments of mortality, complications, readmissions, lengths of stay, and small area variations in activity.

In the UK the NHS reforms depend heavily on accurate administrative information, while the 1990 white paper also called for medical audit based on accurate and complete case note data.¹⁰ Increasingly, as the minimum data set becomes the medium for billing purchasers, providers that fail to complete discharge abstracts in a timely and accurate fashion, are likely not to receive the appropriate payment. If such administrative

CASPE Research, 22
Palace Court, London
W2 4HU

Robert Cleary, research
manager
James Coles, associate
director
Iden Wickings, director

Department of
Obstetrics and
Gynaecology, St
Mary's Hospital
Medical School,
London W2 1PG
Richard Beard, professor

Surgical Audit Unit,
Royal College of
Surgeons of England,
London WC2A 3PN
Brendan Devlin,
chairman

Royal College of
Physicians Research
Unit, London NW1
4LE
Anthony Hopkins,
director

Commission on
Professional and
Hospital Activities,
Ann Arbor, Michigan,
United States
Dale Schumacher, senior
medical advisor

Correspondence to:
Dr Cleary

Accepted for publication
10 February 1994

pressures lead to increased detail and accuracy in routine data then comparative analyses of the resulting large scale databases will be better able to answer not only managerial but also clinical and research questions.

In suggesting that a database built on routinely collected discharge data can address questions relating to quality of care it is important to define the likely limitations of such data. Schumacher and Coles offered four categories of analysis relevant to health care, each offering different levels of aggregation and clinical detail: policy studies, management analysis, clinical analysis (for example, medical audit), and clinical research (for example, randomised clinical trials) (D N Schumacher, J M Coles, sixth international working conference, patient classification systems – Europe, St Etienne, France, 1990). Discharge abstract databases, offering comparisons between, for example, provider units, would probably be most useful in the middle ground of management and clinical analysis. In the US, the Commission on Professional and Hospital Activities (CPHA) uses a set of “quality indicators” to compare the incidence of a range of clinical complications and adverse events (largely recorded within the diagnostic coding) among matched hospitals contributing to its database. In doing so, it recognises that a relatively high incidence is not necessarily associated with poor quality care and possibly reflects uncontrolled differences in case mix. Thus indicators of this kind are best considered as flags that call attention to areas that might benefit from more specific, local audits or detailed prospective research.^{11 12}

The particular information held within a discharge abstract database, limited by compliance with a standard minimum data set, can mean that the specific items required to address a given question may not be present and proxies must be sought. Within a multi-dimensional concept of quality,^{13 17} some dimensions may be poorly addressed by the database approach. However, it is in the standardisation of the data set that the real value of discharge abstract databases lie. Although the possibility must be remembered that some local variation in the definition of terms may exist, the standard allows comparisons to be made, and norms derived, with reference to substantial populations. For those seeking to address quality of care questions such databases also have the advantage of being a relatively low cost resource. If accurate clinical information can be captured, albeit for primarily administrative purposes, then it makes sense for clinicians, researchers, and audit staff to make maximum, but well judged, use of the resulting database.

This study set about building a small comparative database of discharge abstracts in the UK with two aims: firstly, to assess the opportunities for improving on current standards of data quality (particularly diagnostic data) by a range of practical enhancements to the process of data capture and, secondly, to

demonstrate one method by which such a database might be used in monitoring quality.

Methods

About 2000 discharges were abstracted at each of three hospitals. Discharges from general medicine and general surgery were sampled at two units, and those for maternity cases were sampled at the third. Additional, smaller, obstetric samples were also taken at two other maternity units. Table 1 gives details of all five hospitals, together with the number of episodes contributed by each to the database.

At each of the three sites a research abstractor was appointed to record the provider minimum dataset for the sample of discharges (the abstractor based at hospital OB-1 was also responsible for data collection at the two other obstetric sites). In each case the abstractor was an experienced nurse, who received two weeks’ intensive training in clinical coding by CPHA nosologists. The training emphasised the depth of diagnostic and procedural information that may be collected from the clinical record using the relevant coding and classification systems. The instructions to the abstractors emphasised accuracy and completeness of coding over speed and productivity and specified a process of data capture that differs from the norm in the UK in two respects: most importantly, the abstractors worked directly with the clinical notes and not from a discharge summary, coding form, or GP’s letter and, secondly, the abstractors used a clinical encoding software package (Accucode II) to generate the diagnosis and procedure codes. This software generated the required code from the medical terminology entered and the answers to a series of multiple choice questions designed to guide the coder in a consistent manner to the code that represents the maximum level of detail that can be supported by both the coding system and the notes.

Data collection covered discharges from two discrete periods. The first entailed obtaining the notes of a small sample of 1990 patient episodes (about 500 notes) from each of the five sites. These retrospective samples allowed the abstraction by the abstractor and the hospitals’ routinely collected data to be compared when the hospitals’ system could not have been influenced by the presence of the study on site. The second period of data collection (the “concurrent” phase) ran from April to October 1991, during which time the abstractors were based at the study hospitals.

The design called for a random sample of discharges to be drawn from a list of all discharges from the relevant specialties, provided by the hospitals’ patient administration system. However, at hospital MS-1, this proved impossible as the hospital was unable to provide a list of discharges that identified the specialty. In this instance case notes were selected pseudo-randomly from those retrieved from relevant outpatient clinics. From each set of notes the latest discharge from

Table 1 Study hospitals

Hospital	Description	Specialties sampled	No of discharges abstracted
MS-1	District general hospital in North of England, about 28 000 discharges/year	General medicine General surgery (including urology)	1799
MS-2	London teaching hospital, about 28 000 discharges/year	General medicine* General surgery (including urology)	2126
OB-1	London teaching hospital, about 3000 deliveries/year	Obstetrics	2082
OB-2	District general hospital, about 2400 deliveries/year	Obstetrics	650
OB-3	District general hospital, about 5000 deliveries/year	Obstetrics	691

*Includes a group of specialties designed to match the broader definition used elsewhere, in particular at MS-1.

general medicine, general surgery, or urology during 1990 (if any) was included in the sample. In addition, about 5% of the sample was drawn from notes of deceased patients. In the concurrent phase at hospital MS-1 it still proved impossible to obtain a suitable and timely sampling frame, and thus a truly random sample was again impossible. Instead, about 1300 sets of notes were selected pseudo-randomly from those being processed on discharge by the offices of the relevant consultants. Despite efforts to ensure a representative sample both prostatectomies and routine cystoscopies are known to have been under-represented in the available cases.

At hospital MS-2 discharge lists supplied by the patient administration system were used to take a random sample of one in seven discharges from the relevant specialties over the six months from April to September 1990. In about 8% of cases the notes were unavailable and these cases were excluded from the coding process. In the concurrent phase similar methods were used to sample one in two discharges from the same specialties. However, with the shorter lag between discharge and coding implied by the concurrent phase, the proportion of cases for which notes could be accessed fell to around 60%. Though we were unable to identify any systematic features of the notes that were unavailable to us, the possibility of a bias of some kind must be borne in mind.

At the units supplying obstetric data comprehensive lists of deliveries were available, allowing consecutive cases to be reviewed at hospitals OB-1 and OB-2, with alternate cases being sampled at hospital OB-3. The absence of any possible bias in the obstetric sample allowed a greater range of more detailed comparisons between units within this specialty. For this reason, and owing to other characteristics of obstetric care (such as the concept of "primary diagnosis" being of less relevance than in medicine or surgery), where appropriate we restricted our analysis to hospitals MS-1 and MS-2 and reported the obstetric results separately.

DATA QUALITY

To compare the output of the abstractors with that of the hospitals' own centralised data collection system, the abstractors' abstracts from the retrospective phase were matched

with the equivalent records held on the hospitals' information systems. Records were matched on the basis of hospital number and admission date; at hospital MS-2 matches could be obtained for only around 80% of admissions. This seems to reflect, in part at least, errors in the recording of the admission date. If the criterion for a match was relaxed to include admission dates within five days of the date recorded by the abstractor, about half the residual could be "matched." In reporting the comparisons between data obtained by the abstractors and those from matched hospital records only those cases with an exact match were considered.

Depth of diagnostic coding

In most episodes of care it is likely that, in addition to a primary diagnosis, one or more secondary diagnostic factors should be included in a complete abstract. These might represent particular complications or co-morbidities, relevant history, external causes, or other factors influencing health status. As a measure of the extent to which such diagnostic detail was captured by the abstractors and by the hospital information systems the number of diagnosis codes per episode was calculated for each of these sources at both hospitals.

Primary diagnosis

A random sample ($n = 501$) of the matched, coded records from the abstractors and hospital systems at both hospitals was examined by a coding expert from CPHA for differences in the assignment of the primary diagnosis. A further random subsample of these records ($n = 38$) was additionally coded, for validation purposes, by that expert.

Clinical review of ambiguous cases

Twenty cases were selected from those for which coding was available from all three sources (abstractor, hospital system, and CPHA). We wished to determine whose clinical coding would be considered the most accurate in instances in which each of the three sources produced significantly different abstracts. Thus, the basis for selection was the presence of clinically important discrepancies in the primary or secondary diagnosis coding provided by the different sources. The selection was also designed to include equal numbers of medical and surgical cases from each of hospitals MS-1 and MS-2. Two consultant physicians were each asked which of three brief narrative descriptions, based on the three sets of diagnosis and procedure codes and identified by code letters only, came closest to their interpretation of the case notes. The option was given to reject all three descriptions if it was felt that none provided an acceptable account. Case notes for two of the episodes were unavailable for examination, and these cases were excluded from the review.

Reliability

The reliability with which clinical information was abstracted was assessed with a random

sample of 36 medical and surgical cases (17 from hospital MS-1 and 19 from hospital MS-2) that had been included in the concurrent phase, which were reviewed by all three abstractors.

QUALITY INDICATORS

Hospital wide indicators

In order to demonstrate a potential quality assurance application of comparative databases the abstractors' discharge abstracts from hospitals MS-1 and MS-2 were analysed with a selection of the 60 "hospital wide" (that is, not specialty specific) quality indicators developed by the CPHA to report the incidence of a variety of complications and adverse events (each indicator has a definition and detailed coding specification, details of which are available from the project team). The subset of indicators used in this study, selected after discussions with clinicians in the UK, focused on unambiguous, and relatively rare, adverse events (box). To establish whether our efforts to improve the quality of the discharge abstracts would influence the results of the quality indicator analysis, the same indicators were used to analyse the matched records previously obtained from the hospitals' own information systems. The incidence of the indicators was also calculated for 1000 record samples from each of two American hospitals chosen to match the characteristics of hospitals MS-1 and MS-2. These were a district general type hospital in Michigan (US-1) and a large university hospital in Maryland (US-2) respectively. Finally, the CPHA's database was also used to provide overall US projections for the indicator frequencies.

Quality indicators used in study

Hospital wide indicators

Patients with transfusion reaction
Central venous line and septicemia
Cardiac arrest, not principal diagnosis
Cardiac arrest, expired, not principal diagnosis
Respiratory arrest, not principal diagnosis
Respiratory arrest, expired, not principal diagnosis
Patients with decubitus ulcer
Overdose of medication
Patients with principal diagnosis of sign/symptom

Obstetric specific indicators

Admission to special care baby unit
Episiotomy
Obstetric trauma
Induction of labour

Obstetric specific indicators

A range of indicators specific to obstetrics were defined in discussion with clinicians in the UK (mostly from hospital OB-1), as the UK dataset in this specialty is particularly rich. As with the hospital wide indicators, a selection of obstetric indicators (box) was used to analyse not only the abstracts provided by the

abstractors but also the matched records obtained from the hospitals' own information systems.

Results

DATA QUALITY

Depth of diagnostic coding

Table 2 shows a substantial difference in the number of diagnosis codes (including the primary code) per episode at each site: threefold at hospital MS-1 and twofold at hospital MS-2. However, these figures include a substantial proportion of hospital data records with no diagnosis coding. To obtain an accurate indication of the level of secondary coding the number of codes per episode was calculated for those episodes that had at least a primary diagnosis. Even when uncoded records were excluded from the analysis, considerably more diagnostic information was extracted from the notes by the research abstractors than was recorded by the hospitals' own systems (table 2, final row). These figures suggest that the content of the case notes is not the limiting factor in recording diagnostic information. In the case of the hospital information system at hospital MS-1 the large number of uncoded consultant episodes may be associated with the high proportion of admissions recorded as comprising more than one episode. Of the 322 admissions represented by the 446 consultant episodes, 297(92%) had at least one episode coded. At hospital MS-2, moving the analysis of uncoded records to the level of admissions made no difference to the observed level of completion: of 409 admissions, 366(89%) had at least one episode coded.

Primary diagnosis

The comparisons between the primary diagnoses assigned by the abstractors and the equivalent diagnoses held on 501 matched hospital information system records were categorised as follows: 255 records (51%) had identical codes and a further 195(39%) had related diagnoses represented by different codes. Most commonly this occurred when a symptom was recorded by the hospital system instead of a definitive diagnosis (for example, benign prostatic hyper trophy versus bladder neck obstruction) or when a specific diagnosis was contrasted with one that is less specific (for example acute idiopathic pericarditis versus pericardial disease not otherwise specified). The final group, comprising 51(10%) records, represent the cases where major differences occurred in the identification of primary diagnoses. The coding of potentially ambiguous cases is considered further below. Of the 38 episodes additionally coded by the CPHA expert, in 29(76%) the abstractor and the expert agreed on the primary diagnosis, a level of agreement which was considered good, given that identical codes were required for agreement, even closely related diagnoses being considered discrepant.

Clinical review of ambiguous cases

Table 3 shows the results of the review of the 18 sets of discrepant coding narratives by two

Table 2 Depth of diagnostic coding for research abstractors versus hospital information systems at two hospitals

	Hospital MS-1		Hospital MS-2	
	Hospital information system	Research abstractor	Hospital information system	Research abstractor
Consultant episodes*	446	474	447	545
Mean (median) No of diagnosis codes per episode	1.0(1)	2.9(3)	1.6(1)	2.9(3)
Percentage of coded episodes (that is, at least 1 diagnosis)	67	100	89	99
Mean (median) No of codes per coded episode	1.5(1)	2.9(3)	1.8(1)	2.9(3)

*Cases were matched on an admission basis. As each admission may consist of one or more consultant episodes and different coders may disagree on the division, the matched samples need not contain equal numbers of episodes.

Table 3 Agreement with 18 discrepant coding narratives by two consultant physicians

Sources of narrative selected as most accurate	No (%) of narratives	
	Consultant A	Consultant B
Research abstractor	10 (55.5)	7 (38.9)
CPHA expert	5 (27.8)	5 (27.8)
Hospital information system	2 (11.1)	6 (33.3)
Narratives ungraded	1 (5.5)	0

consultant physicians. Overall, the abstractor's narrative was selected as the most accurate on a total of 17 (47.2%) occasions. Surprisingly, though this was better than the hospital system (first choice eight times, 22.2%), it was also an improvement on the rate achieved by the CPHA coder (10 occasions, 27.8%). These figures must be seen in the context of the 33.3% agreement rate that would be expected by chance; while there is some evidence that consultant A favoured the abstractors' abstracts over the other two sources, the observed selections fell short of statistical significance ($p = 0.056$, $\chi^2 = 5.76$, $df = 2$). Consultant B was clearly more equivocal, with a distribution of selections that is very close to that expected by chance ($p = 0.85$, $\chi^2 = 0.33$, $df = 2$).

The results should also be seen in the light of the level of agreement between the clinical reviewers. The fact that they agreed on the source of the most accurate abstract in only 11 of the 18 cases suggests that there was indeed some genuine ambiguity within the selected case notes. In such cases, it may be that a detailed knowledge of the organisation of case notes, in addition to the availability of time and skill for reviewing, becomes particularly important.

Reliability

The 36 cases reviewed by all three abstractors were used to examine the level of agreement between coders at two levels: exact match on primary diagnosis (all three abstractors agreed in 25 (69%) cases) and the assignment of a diagnosis related group (all three abstractors agreed in 26 (72%) cases). Diagnosis related groups are of course influenced by the primary diagnosis, but they can also reflect complications, co-morbidities, and operative procedures. Though two cases with different primary diagnoses may have the same diagnosis related group, the latter level of

analysis (or the comparable healthcare resource groups) may be more appropriate for many potential uses of a discharge abstract database.

QUALITY INDICATORS

Hospital wide indicators

Table 4 shows the incidence of eight illustrative indicators as observed at hospitals MS-1 and MS-2. For each site two frequencies are given for records abstracted by the abstractors. These refer to the retrospective sample (for which matched hospital abstracts were obtained) and the later concurrent phase of data collection. Some of the indicators (for example, patients with transfusion reaction) are shown with zero frequency. Others (for example, patients with signs and symptoms as a primary diagnosis) show a consistent occurrence. Reasonably similar frequencies were observed by the abstractors across the two phases of the project and at the two hospitals. More detailed quantitative comparisons based on these samples are difficult, given the problems in obtaining notes and the potential for differences in case mix. However, the results for the abstractors show that the information required to support the use of such of indicator can be retrieved from the notes.

The incidence identified by the abstractors can be contrasted with the indicator frequencies reported from the hospitals' own information systems. These data represent the observed frequencies for the sample of records matched with the abstracts coded by the abstractors in the retrospective phase. Hence the figures shown in the first two columns of the table for hospitals MS-1 and MS-2 should be directly comparable. With the exception of the quality indicator, recording the use of signs and symptoms, none of the indicators had a recorded incidence other than zero, at either of the hospitals, when the hospital data were used. While consistency of this result across indicators is notable, the limitations of the sample sizes must be considered. For hospital MS-1 an observed incidence of zero in our sample for the hospital information system would be consistent with actual rate as high as 1%. The result relating to the use of signs and symptoms as primary diagnoses can be attributed to one of two alternative explanations: either about 10% of activity is accounted for by episodes in which no definite diagnosis could be made or this definitive diagnosis could not be deduced from the notes whereas the signs and symptoms that had led to the admission could.

The observed frequencies of the eight indicators are also given for three sources of US data. US-1 refers to 1001 discharges from a community hospital in Michigan, matched with hospital MS-1. US-2 refers to a sample (of identical size) of discharges from a university hospital in Maryland, matched with hospital MS-2. The national projection is based on the CPHA 700 hospital database. Encouragingly, these data show frequencies of similar order to those recorded by the

Table 4 Number (percentage) of selected hospital wide quality indicators, as recorded by research abstractor and hospital information system at two UK hospitals and for three US comparators

Quality indicator	Hospital MS-1			Hospital MS-2			US hospitals		
	Hospital information system (n = 446)	Research abstractor		Hospital information system (n = 577)	Research abstractor		US-1 (community) (n = 1001)	US-2 (teaching) (n = 1001)	National projection (n = 34 million)
		Retrospective phase (n = 474)	Concurrent phase (n = 1304)		Retrospective phase (n = 545)	Concurrent phase (n = 1585)			
Patients with transfusion reaction	0	0	0	0	0	0	0		
Central venous line and septicaemia	0	0	0	0	0	0	0		
Cardiac arrest, not principal diagnosis	0	9 (1.9)	15 (1.1)	0	3 (0.6)	23 (1.4)	6 (0.6)	9 (0.9)	(0.8)
Cardiac arrest, expired, not principal diagnosis	0	5 (55.6)	9 (60.0)	0	2 (33.3)	11 (47.8)	5 (83.3)	4 (44.4)	(0.2)
Respiratory arrest, not principal diagnosis	0	2 (0.4)	4 (0.3)	0	0	12 (0.8)	2 (0.2)	8 (0.8)	(0.9)
Respiratory arrest, expired, not principal diagnosis	0	2 (100.0)	1 (25.0)	0	0	1 (8.3)	0	2 (25.0)	(0.9)
Patients with decubitus ulcer	0	0	0	0	2 (0.4)	14 (0.9)	10 (1.0)	6 (0.6)	(1.0)
Overdose of medication	0	1 (0.2)	3 (0.2)	0	1 (0.2)	10 (0.6)	1 (0.1)	0	(0.6)
Patients with principal diagnosis of sign/symptom	45 (10.1)	42 (8.9)	156 (12.0)	35 (6.1)	54 (9.9)	190 (12.0)	53 (5.3)	32 (3.2)	(0.1)

abstractors. Once again, this provides support for the contention that case notes in the UK are capable of supporting this kind of comparative analysis, and emphasises the inadequacy of the normal abstraction process as currently practised.

Obstetric specific indicators

Table 5 gives the results for a selection of the obstetric quality indicators, defined in discussion with UK clinicians. Again, the frequencies observed from the abstractors' abstracts are compared with the incidence within the matched records from the routine information system. In the case of the three obstetric sites, the routine information system (the St Mary's Maternity Information System (SMMIS) is a departmentally based, online data collection system with semiautomated clinical coding, in contrast to the centralised manual coding departments supporting hospitals MS-1 and MS-2.

As noted previously (R W Beard *et al*, unpublished data), the levels of completeness and accuracy associated with SMMIS data collection were higher than we have reported here for the more centralised systems in medicine and surgery. The results obtained with the obstetric quality indicators reflect this situation, the rates returned by the abstractors and the hospital information systems being mostly comparable and able to distinguish between units. Where discrepancies exist between the two sources, as in the case of the rates of induction of labour a more detailed

analysis has led to the conclusion that systematic, correctable differences in the definition and interpretation of clinical terms tend to underlie the problem (R W Beard *et al*, unpublished data).

Discussion

The current levels of completeness, accuracy, and reliability of computerised discharge abstracts (as stored on hospital information systems) are still a cause for concern and are capable of substantial improvement. Past studies found that the accuracy or completeness of data, or both, ranged from 60%–90% with certain codings being "totally inexplicable."^{18–20} Although we recognise that great efforts have been made to improve the situation in parallel with the current NHS reforms and that undoubtedly there has been recent progress,^{21–22} our analysis has shown similar results in the medical/surgical study.

The routine abstraction of data at hospitals MS-1 and MS-2 showed major deficiencies in the database drawn from the hospitals' own information systems. Incomplete or inaccurate recording of diagnostic information will generally invalidate the use of such data when addressing questions relating to the quality of care. Even simple applications of the data, such as its use to define a sampling frame for more detailed data collection, are potentially undermined. Though it was not the purpose of this study to examine the underlying causes of poor quality in hospital recording systems, it is clear that some coders are required to work from summarised forms and do not see the whole clinical note. Unless the summary contains full details of secondary diagnoses, including complications and co-morbidities, it is not surprising that the coded details are less than perfect.

One effect of the current deficiencies in routine data collection was disclosed by the use of the simple hospital wide quality indicators. If the hospitals' data are used they give the impression that none of these indicator events occurred. It seems reasonable to conclude that this is not the case. Firstly,

Table 5 Number (percentage) of selected obstetric quality indicators, as recorded by research abstractor and hospital information system at three hospitals

Quality indicator	Hospital OB-1		Hospital OB-2		Hospital OB-3	
	Hospital information system (n = 313)	Research abstractor (n = 313)	Hospital information system (n = 258)	Research abstractor (n = 258)	Hospital information system (n = 321)	Research abstractor (n = 321)
Admission to special care baby unit	13(4.1)	18(5.7)	18(7.0)	19(7.4)	31(9.7)	44(13.7)
Episiotomy	81(25.9)	78(24.9)	47(18.2)	46(17.8)	101(31.5)	101(31.5)
Obstetric trauma	129(41.2)	117(37.4)	110(42.6)	103(39.9)	120(37.4)	112(34.9)
Induction of labour	70(22.4)	31(9.9)	86(33.3)	40(15.5)	141(43.9)	58(18.0)

the occurrence of some of these events can be verified by independent validation or through other reporting systems. Secondly, the occurrence of such events was noted by the abstractors from the same notes, but it depended on the accurate coding of secondary diagnoses. As we have reported, the hospital systems record significantly less information than can be obtained from the notes. Even when using the abstractors' data to calculate the various indicators, it is clear that we are dealing with infrequent events. As such, even quite large samples may leave the results of any comparisons within the bounds of chance variation. If indicators that might be expected to yield higher incidences were used (for example, wound infection rates) there may still be a problem of attributing any observed variation to a specific problem. A recent review of the appropriateness of routine data for use in clinical audit draws together a variety of evidence relating to both the inherent limitations of the International Classification of Diseases for diagnostic coding and problems that arise from its typical use in practice.²³ Concato *et al*, in completing a detailed cohort study of mortality after prostatectomy, illustrated the difficulty in controlling for confounding factors within database analyses.²⁴ The existence of these limitations underline the point that, in the case of questions relating to quality of care, database analysis may be best seen as an approach that can call attention to areas requiring more detailed investigation.

Across the NHS the enhanced need for reliable and valid information is self evident. Chantler identified the need for improved information for management and the need to support medical audit.²⁵ Others have called attention to the need for accurate data to support more detailed understanding of the health system – for example, in defining case mix and severity issues.²⁶ To meet these needs researchers have chosen to use other source documents for more detailed clinical analyses and studies of quality of care.^{27 28} Though there will always be a need for specific studies and accompanying protocols for additional data collection, the lack of a reliable database of basic discharge data has resulted in an unnecessary proliferation of separate and disconnected systems.

However, we believe our study is a cause for optimism in that clinical case notes within the sampled hospitals contained sufficient data to provide a full patient abstract, as defined by the provider minimum dataset (PMDS). In particular, they supported an increased level of coding of diagnosis and procedure codes which, as noted in a recent study in the US has been associated with more general improvements in the accuracy of discharge abstracts as a valid representation of the case notes.²⁹ Our enhancements of the data collection process were sufficient to raise the quality of the discharge abstracts to a level where plausible values were obtained for a range of hospital wide quality indicators.

Naturally, such enhancements may have an

associated cost. The experience of the abstractors in the present study leads us to believe that, on an operational basis, an abstractor working along similar lines would be able to code over 5000 discharges per year. This figure is below that currently expected from coders in the UK, but it is entirely consistent with the productivity that the CPHA expects from abstractors in the US. The St Mary's maternity information system, which seems already to be capable of supporting specialty specific quality indicators, may also offer a way forward. It exemplifies a system for data capture that is integrated within the department in which the process of care takes place. It seems likely that the routine use of the data it holds, both operationally and in monitoring the quality of care, helps in maintaining high standards of data quality. We believe that such investments in data capture will become increasingly attractive as the specificity of discharge abstracts become valuable for billing the purchasers of health care. If clinicians and managers can break the vicious circle which links poor quality routine data with its disuse, then those attempting to address questions relating to the quality of care will benefit.

We thank the managers, clinicians, and staff at the five study hospitals, in particular Drs A. Tanner and C. Paterson, for their help.

- 1 Information Management Group. *Minimum data set model, version 2*. London: Department of Health, 1991.
- 2 Barrie JL, Marsh DR. Quality of data in the Manchester orthopaedic database. *BMJ* 1992;**304**:159–62.
- 3 Lyons C, Gumpert R. Medical audit data: counting is not enough. *BMJ* 1990;**300**:563–6.
- 4 Walshe K, Harrison N, Renshaw M. Assessing the quality of patient data collection by hospital and departmental computer systems. *Health Trends* (in press).
- 5 Institute of Medicine. *Reliability of hospital discharge abstracts*. Washington, DC: National Academy of Sciences, 1977. (Publication IOM-77-01.)
- 6 Hsia DC, Krushat WM, Fagan AB, Tebbutt JA, Kusserow RP. Accuracy of diagnostic coding for medicare patients under the prospective-payment system. *N Engl J Med* 1988;**318**:352–5.
- 7 Grady ML, Schwartz HA. Medical effectiveness research data methods. Agency for Health Care Policy and Research, 1992. Rockville, MD: (AHCPR Publication No 92-0056.)
- 8 Wennberg JE, Roos N, Sola L, Schori A, Jaffe R. Use of claims data to evaluate health care outcomes. Mortality and reoperation following prostatectomy. *JAMA* 1987;**257**:933–6.
- 9 Salive ME, Mayfield MD, Weissman NW. Patient outcomes research teams and the Agency for Health Care Policy and Research. *Health Serv Res* 1990;**25**:697–708.
- 10 Secretaries of State for Health, Wales, Northern Ireland, and Scotland. *Working for patients*. London: Department of Health, 1989. (Cm 555.)
- 11 Collopy BT. Audit activities in Australia. *BMJ* 1991;**303**:1523–5.
- 12 Collopy BT. Developing clinical indicators: the ACHS Care Evaluation Program. *Aust Clin Rev* 1990;**10**:83–5.
- 13 Steffen GE. Quality medical care: a definition. *JAMA* 1988;**260**:56–61.
- 14 Mercy Health Services. *Framework for managing quality*. St Josephs, MO: Sisters of Mercy Health Corporation, 1987.
- 15 Maxwell RJ. Quality assessment in health. *BMJ* 1984;**288**:1470–2.
- 16 Tarlov AR, Ware JE, Greenfield S, Nelson EC, Perrin E, Zubakoff M. The medical outcomes study: the application of methods for monitoring the results of medical care. *JAMA* 1989;**262**:925–30.
- 17 Hopkins A. *Measuring the quality of medical care*. London: Royal College of Physicians of London, 1990.
- 18 Whates PD, Birgzzalis AR, Irvine M. Accuracy of Hospital Activity Analysis codes. *BMJ* 1982;**284**:1857–8.
- 19 Rees JL. Accuracy of Hospital Activity Analysis data in estimating the incidence of proximal femoral fractures. *BMJ* 1982;**284**:1856–7.
- 20 Sunderland R. Inaccurate coding corrupts medical information. *Arch Dis Child* 1985;**60**:593–4.
- 21 Information Management Group. *A national thesaurus of*

- clinical terms in Read codes*. London: Department of Health, 1992.
- 22 Information Management Group. *Setting standards for coding clinical data*. London: Department of Health, 1993.
 - 23 McKee M. Routine data: a resource for clinical audit. *Quality in Health Care* 1993;2:104–11.
 - 24 Concato J, Horowitz RI, Feinstein AR, Elmore JG, Schiff SF. Problems of co-morbidity in mortality following prostatectomy. *JAMA* 1992;267:1077–82.
 - 25 Chantler C. Management and information. *BMJ* 1992;304:632–5.
 - 26 Bardsely M, Coles J, Jenkins L. *DRGs and health care. The management of case mix*. 2nd ed. London: King Edward's Hospital Fund for London, 1989.
 - 27 Campling EA, Devlin HB, Hoile RW, Lunn JN. *Report of the National Enquiry into Perioperative Deaths*. London: Royal College of Surgeons, 1990.
 - 28 Hopkins A, Costain D, eds. *Measuring the outcomes of medical care*. London: Royal College of Physicians of London, 1990.
 - 29 Hsia DC, Ahern CA, Ritchie BP, Moscoe LM, Krushat WM. Medicare reimbursement accuracy under the prospective payment system, 1985 to 1988. *JAMA* 1992;268:896–9.