

Systematic review of discharge coding accuracy

E.M. Burns¹, E. Rigby¹, R. Mamidanna¹, A. Bottle², P. Aylin², P. Ziprin¹, O.D. Faiz¹

¹Department of Surgery, Imperial College, St Mary's Hospital, Praed Street, W21NY London, UK

²Dr. Foster Unit, Department of Primary Care and Public Health, Imperial College, 1st Floor Jarvis House, 12 Smithfield Street, EC1A 9LA London, UK

Address correspondence to Omar Faiz, E-mail: o.faiz@imperial.ac.uk

ABSTRACT

Introduction Routinely collected data sets are increasingly used for research, financial reimbursement and health service planning. High quality data are necessary for reliable analysis. This study aims to assess the published accuracy of routinely collected data sets in Great Britain.

Methods Systematic searches of the EMBASE, PUBMED, OVID and Cochrane databases were performed from 1989 to present using defined search terms. Included studies were those that compared routinely collected data sets with case or operative note review and those that compared routinely collected data with clinical registries.

Results Thirty-two studies were included. Twenty-five studies compared routinely collected data with case or operation notes. Seven studies compared routinely collected data with clinical registries. The overall median accuracy (routinely collected data sets versus case notes) was 83.2% (IQR: 67.3–92.1%). The median diagnostic accuracy was 80.3% (IQR: 63.3–94.1%) with a median procedure accuracy of 84.2% (IQR: 68.7–88.7%). There was considerable variation in accuracy rates between studies (50.5–97.8%). Since the 2002 introduction of Payment by Results, accuracy has improved in some respects, for example primary diagnoses accuracy has improved from 73.8% (IQR: 59.3–92.1%) to 96.0% (IQR: 89.3–96.3), $P = 0.020$.

Conclusion Accuracy rates are improving. Current levels of reported accuracy suggest that routinely collected data are sufficiently robust to support their use for research and managerial decision-making.

Keywords epidemiology, health services, management and policy

Introduction

Routinely collected data are increasingly used at local, national and international levels for epidemiological studies, clinical research, audit, health resource distribution, and developing health-care policies and funding strategies.

Several national bodies collect data regarding patient hospital attendances recording diagnoses and procedures using the World Health Organization's International Classification of Diseases (ICD)¹ and operative interventions and procedures with Office of Population, Censuses and Surveys (OPCS) classification of interventions and procedures, fourth revision.² Hospital Episode Statistics (HES) record all admissions and (from 2003) outpatient attendances in NHS hospitals in England. Patient Episode Database for Wales (PEDW) and the Scottish Morbidity Record (SMR)

record hospital attendances in Wales and Scotland, respectively.

In 2001, Campbell *et al.*³ conducted a systematic review on accuracy of UK routinely collected data. Accuracy was high overall (84% for diagnostic codes and 97% for procedures). Since this review, there have been changes to

E.M. Burns, Clinical Research Fellow

E. Rigby, Academic Foundation Year 2

R. Mamidanna, Clinical Research Fellow

A. Bottle, Lecturer in Medical Statistics

P. Aylin, Clinical Reader in Epidemiology and Public Health

P. Ziprin, Senior Lecturer and Honorary Consultant Colorectal Surgeon

O.D. Faiz, Senior Lecturer and Honorary Consultant Colorectal Surgeon

coding practices, including the introduction of Payment by Results (PbR) and to OPCS and ICD classifications. PbR is an initiative directing health-care funding based on coding data. A clinical audit programme, carried out in all acute NHS trusts, showed that errors in coding had significant impact on payment accuracy.⁴ Average Health-care Resource Group (HRG) coding error was 9.4% (range: 0.3–52% across trusts), an error of £3.5 million. Although the net financial impact was close to zero, in some cases the local impact was significant. The NHS Operating Framework for 2008–09 calls for a focus on clinical coding in the drive for world-class patient care.⁵

The accuracy of routinely collected data can be assessed against various standards. In this review, the ‘gold standard’ is assumed to be comparison with independent case note review. This requires reliable data within the case notes. Where indicated, coding is compared with other sources such as clinical registry data. Each system is subject to possible inaccuracy as the data quality depends on those inputting data. In addition, registries may not use OPCS or ICD-10 coding systems. Studies that use clinical registry data are considered separately from case note studies.

The primary objective of this study is to identify and review studies investigating the accuracy of hospital episode data. Secondary objective is to investigate factors influencing variation in coding.

Methods

The measurement tool for ‘assessment of multiple systematic reviews’ (AMSTAR), which consists of 11 items for assessing methodological quality of systematic reviews, was employed.⁶

Literature search

We searched PubMed, EMBASE, The Cochrane Database and Ovid to identify studies assessing the accuracy of hospital coding data from Great Britain. Studies published from 1989 to present were included. Using the search term ‘PEDW’ did not yield any further relevant articles. References were hand searched for further relevant articles. Expert knowledge of potential further sources, such as the Audit Commission, was used to ensure comprehensive review of available sources. Papers were assessed using a pre-defined checklist of quality criteria derived from Crombie⁷ and utilized previously by Campbell *et al.*³ The search terms, quality and inclusion criteria are shown in Box 1.

Box 1 Search terms and quality assessment criteria.

Search terms

1. Scottish Morbidity Record, OCD, SMR, OPCS, ICD (MeSH), HES, HAA
2. Classification, nomenclature (includes vocabulary controlled) (MeSH), Medical records (MeSH), Medical records, computerised (MeSH), Medical Record Linkage (MeSH), Registries (MeSH), Forms and record control, clinical coding.
3. Accuracy (Ti/Ab), Quality (Ti/Ab)
4. Limit year 1989 to present
5. Great Britain
6. 5 and 6
7. 1 and 3
8. 2 and 3
9. 1 and 2 and 3
10. 6 and (7 or 8 or 9)

Inclusion criteria

1. Compare routinely collected hospital coding data with independent review of hospital notes or discharge summaries
2. Examine ICD and/or OPCS codes
3. Measure data quality against published standards and rules
4. Be based in Great Britain
5. Be published in the English language
6. Be published after 1989
7. Have identifiable accuracy rates

Quality Assessment

1. Random sampling of episodes. This was coded as ‘yes’ if random sampling was explicitly stated or all episodes from a defined time period were obtained; ‘no’ if sampling was mentioned, but not random and ‘unclear’ when the sampling strategy was not outlined.
2. At least 90% of episodes sampled were available for analysis. This was coded as ‘yes’ if the percentage was >90%; ‘no’ if the percentage was <90% and ‘unclear’ when the percentage was not recorded or able to be calculated from the data.
3. Trained coders were utilised. This was coded as ‘yes’ when coders training or experience was specifically mentioned; ‘no’ when coders were stated as clinicians or untrained and ‘unclear’ when the training of coders was not mentioned.
4. Inter- and intra-coder reliability rates were reported. This was coded as ‘yes’ when rates were recorded; ‘no’ when no record of reliability rates was made and ‘unclear’ when reliability was discussed but not explicitly stated.
5. Awareness of codes at time of discharge. This was coded as ‘no - unaware’ when coders were blinded to the original coding of a procedure or diagnosis; ‘yes - aware’ when coders were aware of the original diagnoses when recoding case notes or discharge summaries or ‘unclear’ when awareness of coders to previous coding was not noted.

Table 1 Assessment of quality of studies examining data accuracy of routinely collected data in comparison to case note review

First Author	Year	Data source	Random sampling	90% sampled available	Trained coders	Coder reliability	Coder awareness of codes	Definition of accuracy
Sellar <i>et al.</i> ²⁶	1990	Registry and case note	No	Yes	Unclear	No	Yes, aware	Unclear
Smith <i>et al.</i> ²⁷	1991	Case note review	Yes	Yes	No	No	Unclear	Unclear
Yeoh and Davies <i>et al.</i> ²⁸	1993	Case note review	Yes	No	Unclear	Yes	No, unaware	Unclear
Panayiotou ²¹	1993	Case note review	Unclear	Yes	Yes	No	Yes, aware	Three digit
Cleary <i>et al.</i> ²⁹	1994	Case note review	Unclear	Unclear	Yes	Yes	Unclear	Four digit
Drennan ³⁹	1994	Case note review	Yes	Yes	Yes	No	No, unaware	Unclear
Gibson and Bridgman ¹²	1998	Case note review	Yes	No	Unclear	No	Unclear	Four digit
Dixon <i>et al.</i> ¹⁰	1998	Case note review	Yes	Yes	Yes	Yes	Unclear	Four digit
Kirkman <i>et al.</i> ¹⁵	2009	Discharge summary	Yes	Unclear	Unclear	No	Unclear	Four digit
Reddy-Kolanu and Hogg ²⁴	2009	Case note review	Yes	Yes	Yes	No	Unclear	Unclear
Nouraei <i>et al.</i> ²⁰	2009	Case note review	Yes	Yes	Yes	Unclear	Unclear	Four digit
Mitra <i>et al.</i> ¹⁸	2009	Case note review	Yes	Unclear	Yes	No	Unclear	Four digit
Beckley <i>et al.</i> ³¹	2010	Case note review	Yes	Unclear	Yes	No	Unclear	Unclear
Audit Commission ³⁰	2010	Case note review	Yes	Unclear	Yes	Unclear	Unclear	Four digit
Murchison <i>et al.</i> ¹⁹	1991	Case note review	No	Yes	Unclear	No	Unclear	Unclear
Park <i>et al.</i> ²²	1992	Case note review	No	Yes	No	No	Unclear	Unclear
McGonigal <i>et al.</i> ¹⁷	1992	Case note review	No	Yes	No	No	Yes, aware	Four digit
Pears <i>et al.</i> ²³	1992	Case note review	Unclear	No	Unclear	Unclear	No, unaware	Four digit
Samy <i>et al.</i> ²⁵	1994	Case note review	Yes	Unclear	Unclear	Unclear	Unclear	Unclear
Dornan <i>et al.</i> ¹¹	1995	Case note review	Yes	Yes	Yes	No	Yes, aware	Unclear
Harley and Jones ¹³	1996	Case note review	Yes	Yes	Yes	No	Unclear	Three digit
Davenport <i>et al.</i> ⁹	1996	Case note review and local registry	No	Yes	Unclear	No	Yes, aware	Unclear
Kohli <i>et al.</i> ¹⁶	2009	Case note review	Yes	Yes	Unclear	No	Yes, aware	Four digit
Hasan <i>et al.</i> ¹⁴	1995	Case note review	Yes	Yes	Yes	No	Unclear	Four digit
Colville <i>et al.</i> ⁸	2000	Operation note review	Yes	Yes	No	No	No, unaware	Four digit

Studies from the electronic searches were reviewed independently by E.B. and E.R. Discrepancies between selected papers were assessed by R.M. for inclusion and agreed through consensus. All papers assessing accuracy of hospital coding data were included and no restrictions were made on the type of study.

Reported accuracy refers to the primary diagnosis and main procedure code. Accuracy is defined as the percentage agreement between coding allocated through independent assessment of hospital notes or discharge summaries and that recorded on the routinely collected data set. The overall diagnosis and procedure accuracies were calculated where applicable. In those studies that assessed the accuracy of both the procedure and diagnosis, if stated in the paper, the overall accuracy was used to contribute to calculation of the median overall accuracy of the studies. If not stated in the paper, diagnostic and procedure accuracies were considered separately. Some studies report three- or four-level accuracy. The accuracy level reported is that described by

the authors of the individual studies as stated in Table 1. The clinicians' diagnosis at discharge was the standard against which accuracy was measured.

Results

Sixty-nine potential studies were identified by the searches. Of these, 37 studies were excluded. Figure 1 shows the reason for excluding studies. Of the 32 included studies, 25 studies compared the accuracy of routinely collected data with case or operation notes^{8–31} and seven studies contrasted routinely collected data with clinical registry data.^{32–38} Tables 2 and 3 summarize the details of the included studies that used case note review and registry data, respectively. Of the papers that compared routinely collected data accuracy with case note review, 14 papers (56%) used English data sets^{10,12,15,18,20,21,24,26–31}, 9 (37.5%) examined Scottish data^{9,11,13,16,17,19,22,23,25} and two studies used Welsh data.^{8,14} Twenty of these papers assessed the accuracy of

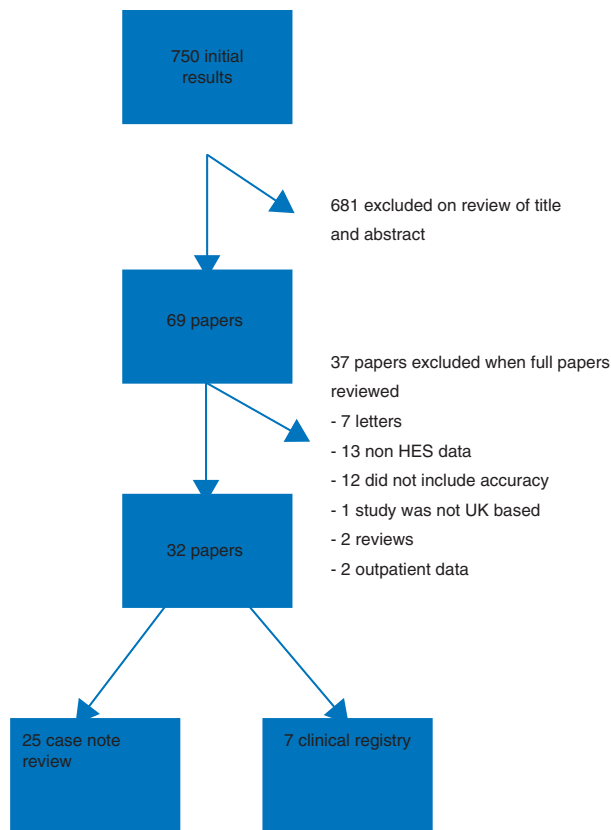


Fig. 1 Schematic of inclusion following literature search.

diagnostic coding^{8–12,14–17,19–23,25,26,28,29} and 9 papers assessed the accuracy of procedure coding^{8,10,13,18,20,24,27,39}. The majority of studies that assessed diagnostic coding accuracy used ICD-9 (11 studies) exclusively. Four studies examined ICD-10 and three studies with long study periods used a combination of ICD-9 and ICD-8. A version of the OPCS-4 coding system was used in seven of the nine studies that examined procedure coding. The remaining two studies used OPCS-3 or an unspecified version of OPCS system.

Papers comparing routinely collected data with case note review

Study quality

The studies varied in size of included admissions from 34 to 17 959 admissions with a median of 298 admissions. Table 1 summarizes the quality assessment for each of these studies. Seventeen studies stated that their samples were random. Sixteen studies assessed >90% of the case notes selected for sampling. Ten studies stated that trained coders were used and three studies assessed inter-coder reliability. Six studies stated that the coders performing case note

review were blinded to the original codes. Table 1 states the level of accuracy assumed for each study.

Accuracy

The overall median accuracy was 83.2% (IQR: 67.3–92.1%). The median diagnostic accuracy was 80.3% (IQR: 63.3–94.1%) with a median procedure accuracy of 84.2% (IQR: 68.7–88.7%).

When we compared those studies that included data prior to the introduction of PBR (2004) and those afterwards, there were no differences in overall coding accuracy [pre-PbR 77.0% (IQR: 66.2–89.0%) versus post-PbR 86.1% (IQR: 73.1–96.1%), $P = 0.207$] or the accuracy of procedure codes ($P = 0.602$) but the accuracy of the primary diagnosis improved [73.8% (IQR: 59.3–92.1%) versus 96.0% (IQR: 89.3–96.2%), $P = 0.020$]. There was no difference in overall accuracy between multiple hospital and single site data sets ($P = 0.252$). When Scottish studies were compared with those assessing English data, there were no differences in overall, procedure or diagnosis accuracy ($P = 0.292$, $P = 0.245$ and $P = 0.742$, respectively).

Those studies that used random sampling for case selection had lower median accuracy [random accuracy 83.1% (IQR: 68.0–88.2%) versus non-random 93.7% (IQR: 90.3–95.0%), $P = 0.033$].

Papers comparing routinely collected data with clinical registry data

Seven studies compared routinely collected data with clinical registries.^{32–38} Five studies compared HES data with national registry data.^{32,33,36–38} Three studies compared number of procedures and mortality against surgical society clinical registries.^{36–38}

A further study examined *Clostridium difficile* rates reported on HES database against those reported to the Health Protection Agency (HPA).³² Reporting cases of *C. difficile* to the HPA is mandatory. Mukherjee *et al.*³³ compared rates of ovarian neoplasms against a local registry and histopathology data set. Two further Scottish studies compared SMR data against local registries.^{34,35} Table 3 summarizes these studies and shows the number of procedures recorded on the registries versus administrative datasets.

HES data recorded twice as many procedures as the National Vascular disease (NVD) registry (HES $n = 16\,923$ and NVD $n = 8462$) with slightly higher death rates recorded on HES (HES, 18% and NVD, 15%).³⁷ Garout *et al.*³⁸ found a higher number of colorectal procedures reported on HES than on the Association of Coloproctology of Great Britain and Ireland (ACPGBI)

Table 2 Summary of included studies examining data accuracy of routinely collected data in comparison to case note review

Country	Study	Year	Diagnosis/procedure included	Study dates	Coding system	Number of cases sampled	Setting	Data accuracy
England	Sellar <i>et al.</i> ²⁶	1990	Deliberate self-poisoning	1980–1985	ICD-8, ICD-9	488	Single hospital	Diagnosis, 95.7%
England	Smith <i>et al.</i> ²⁷	1991	Joint replacements	1988	OPCS3	139	3 hospitals	Procedure, 85.0%
England	Yeoh and Davies ²⁸	1993	Paediatric diagnoses	1990, 1991	ICD	37 1990, 117 1991	Single acute hospital	Diagnosis, 54.1%, 1990 Diagnosis, 84.6%, 1991
England	Panayiotou ²¹	1993	Cerebrovascular disease	Unspecified	ICD-9	117	Single acute hospital	Diagnosis, 76.0%
England	Cleary <i>et al.</i> ²⁹	1994	All general medicine and general surgery diagnoses	1990–1991	ICD	501	2 acute hospitals	Diagnosis, 51.0%
England	Drennan ³⁹	1994	Urology, cardiothoracics, cardiology, general surgery	1990–1991	OPCS4.2 ICD-9	2044	4 acute hospitals	Diagnosis, 68.0% Procedure 83.0%
England	Gibson and Bridgman ¹²	1998	General surgery diagnosis	1995	ICD-10	298	Single acute hospital	Diagnosis 71.0%
England	Dixon <i>et al.</i> ¹⁰	1998	All	1991–1993	OPCS4, ICD-9	Diagnosis, 1252; procedure, 416	2 hospitals	Diagnosis, 50.5% Procedure, 65.9%
England	Kirkman <i>et al.</i> ¹⁵	2009	Haemorrhagic stroke	2002–2007	ICD-10	ICH 978 SAH 1169	4 acute hospitals	Diagnosis, ICH, 95.9% Diagnosis, SAH 96.1%
England	Reddy-Kolanu and Hogg ²⁴	2009	ENT procedures	2008	OPCS4	79	Hospital day surgery unit	Procedure, 69.6%
England	Nouraei <i>et al.</i> ²⁰	2009	Otolaryngology procedures	2007–2008	OPCS4, ICD-9	1250	Single acute hospital	Diagnosis, 96.2% Procedure, 85.1%
England	Mitra <i>et al.</i> ¹⁸	2009	Head and neck surgery procedures	2006	OPCS	34	Single acute hospital	Procedure, 52.6%
England	Beckley <i>et al.</i> ³¹	2009	Urological procedures	2007	ICD-10, OPCS4	500	Single acute hospital	Procedure, 83.4%
England	Audit Commission ³⁰	2010	All diagnoses	2009–2010	ICD-10, OPCS4	Unknown	Multiple hospitals	Overall, 87.0% Diagnosis, 87.0% Procedure, 90.0%
Scotland	Murchison <i>et al.</i> ¹⁹	1991	Inflammatory bowel disease	1968–1983	ICD-8, ICD-9	255	All NHS hospitals in Scotland	Overall, 93.7% Crohn's, 95.5% Ulcerative colitis, 91.0%
Scotland	Park <i>et al.</i> ²²	1992	Wilson's disease	1974–1989	ICD-8, ICD-9	40	All Scotland	Diagnosis, 87.50%
Scotland	Kohli <i>et al.</i> ¹⁶	1992	Gastrointestinal Diagnosis, co-existing Arthritis	1987	ICD-9	778	Multiple hospitals	Diagnosis, 73.6%
Scotland	McGonigal <i>et al.</i> ¹⁷	1992	Dementia	1974–1988	ICD	196	Single hospital	Diagnosis, 93%
Scotland	Pears <i>et al.</i> ²³	1992	Paediatric and general medical diagnoses	Unspecified	ICD-9	52 paediatric 100 medical	Single hospital	Paediatric diagnosis, 67.0% Medical diagnosis, 54.0%

Scotland	Samy <i>et al.</i> ²⁵	1994	Abdominal aortic aneurysm diagnosis	1979–1991	ICD-9	500	Unclear	Diagnosis, 97.80%
Scotland	Dorman <i>et al.</i> ¹¹	1995	Upper gastrointestinal diagnoses	1989–1991	ICD-9	3447	Single hospital	Diagnosis, 58.40%
Scotland	Harley and Jones ¹³	1996	All diagnoses	1992, 1994	ICD-9, OPCS4	17959	Multiple hospitals	Diagnosis, 89.2%, Procedure, 88.2%
Scotland	Davenport <i>et al.</i> ⁹	1996	Stroke	Unclear	ICD-9	566	Single hospital	Diagnosis, 94.2%
Wales	Hasan <i>et al.</i> ¹⁴	1995	Cerebrovascular disease	1993	ICD-9	166	Single hospital	Diagnosis, 74.0%
Wales	Colville <i>et al.</i> ⁸	2000	Plastic surgery procedures	1998	OPCS4	50	Single hospital	Overall, 78.0% Diagnosis, 62.0% Procedure, 98.0%

ENT, ears nose and throat; SAH, subarachnoid haemorrhage; ICH, intracerebral haemorrhage; ICD, International Classification of Disease; OPCS, Office of Population, Censuses and Surveys (OPCS) classification of interventions and procedures.

colorectal cancer database (HES $n = 7516$ and ACPGBI $n = 6617$) with comparable overall mortality at a national level [HES 418 (5.6%) versus ACPGBI 383 (5.8%), $P = 0.416$].³⁶ Westaby *et al.*, however, found a higher number of reported infant cardiothoracic procedures on the Central Cardiac Audit Database (CCAD) than on the HES (HES, $n = 1745$ and CCAD, $n = 2182$). The reported mortality was lower on HES than on CCAD [HES $n = 74$ (4.2%) versus CCAD $n = 139$ (6.4%)]. However, the two data sets differed in the types of procedures included in the analysis with all procedures included in the CCAD and a limited number included in the HES data analysis. The definition of 30-day mortality differed between data sets, with HES recording only those deaths in hospital and the CCAD including all deaths in and out of hospital. Thus, the comparison was inhibited by different coding systems and difficulty in defining the same procedures and outcomes.

Discussion

Main findings of the study

Data accuracy has been a concern for clinicians, managers and central government.⁴⁰ Steps have been taken to improve quality. The Care Quality Commission mandates yearly audits of individual trust data quality.⁴¹ This study examines the accuracy of administrative data in published literature. Overall accuracy was 83% with procedure accuracy (84.2%) found to be higher than primary diagnosis coding (80.3%). Accuracy of diagnostic coding has improved substantially in recent years.

What is already known on this topic

Implications of data accuracy

Questions should be asked as to whether accuracy of 83%, or 87% as quoted by the Audit commission report³⁰, is reasonable to allow the data to be employed for current purposes. There is no consensus of what is acceptable data accuracy. The ultimate goal would be data accuracy of 100%. A more realistic target may be 98%, the highest data accuracy recorded in the literature.²⁵ Clinician involvement in coding has been proposed to improve accuracy.⁴² Yeoh and Davies²⁸ examined changes in accuracy after clinicians became responsible for coding. Accuracy increased from 54 to 85% over a 1-year period. Though, given such a low initial accuracy, it may be argued that there were serious flaws in early coding, questioning the broader applicability of this research. Nouraei *et al.*²⁰ observed that use of a clinician coding multi-disciplinary team resulted in a change to 24.1% of records and an increase in departmental revenue of

Table 3 Summary of studies included comparing routinely collected data with clinical registries

Country	Study	Year	Diagnosis/procedure considered	Study dates	Coding system	Comparison registry	Measure used	Setting	% Data recorded on HES versus registry
England	Jen <i>et al.</i> ³²	2008	<i>Clostridium Difficile</i> , orthopaedic surgical site infection (SSI)	2004–2005	ICD-10	HPA mandatory reporting registry	Numbers included on each database	Multiple hospitals	<i>Clostridium Difficile</i> HPA 93121 HES 36757 SSI HPA 1191 HES 1045
England	Mukherjee <i>et al.</i> ³³	1991	Ovarian neoplasm	1979–1983	ICD-9	Ovarian tumour Registry and Regional Cancer Registry	Case inclusion	Multiple hospitals	Ovarian tumour registry 685 HAA 611
England	Garout <i>et al.</i> ³⁶	2008	Colorectal cancer	2001–2002	OPCS4	National Clinical Registry	Patient volume and outcome	Multiple hospitals	ACPGBI registry 6, 617 cases HES 7, 516 cases Comparable mortality rates
England	Aylin ³⁷	2007	Vascular procedures	2001–2003	OPCS4	National Clinical Registry	Patient volume and outcome	Multiple hospitals	NVD 8462 cases HES 16 923 cases Comparable mortality
England	Westaby <i>et al.</i> ^{38a}	2007	Cardiac paediatric procedures	2000–2002	OPCS4	National Clinical Registry	Case inclusion	Multiple hospitals	CCAD 1745 HES 2182 Mortality—HES 4.2% versus CCAD 6.4%
Scottish	Raza <i>et al.</i> ³⁴	1999	Vascular procedures	1994	ICD-9, OPCS4	Local vascular database	Operative accuracy	Single hospital	Local vascular database 840 cases ISD 793 cases
Scottish	Milburn <i>et al.</i> ³⁵	2007	General, paediatric and vascular surgery	2003–2004	ICD-10, OPCS4	Local database	Accuracy of diagnosis and procedure coding	Multiple hospitals	Clinically acceptable match; diagnosis 86.9%, procedure 84.0%

HPA, Health Protection Agency; ICD, International Classification of Disease; OPCS, Office of Population, Censuses and Surveys (OPCS) classification of interventions and procedures; ACPGBI, Association of Coloproctology of Great Britain and Ireland; NVD, National Vascular Database; HES, Hospital Episode Statistics; CCAD, Central Cardiac Audit Database; ISD, Information and Statistics Division.

^aDefinition of 30 mortality and breadth of included procedures varied between HES and CCAD.

£443 371. This suggests that clinician involvement may be a cost-effective means of improving data quality and hospital reimbursement. Greater education is needed amongst clinicians.

The majority of studies included in this review defined inaccurate coding as inaccurate four digit coding (Table 1). Both OPCS and ICD-10 use four digit codes to signify procedures and diagnoses, respectively. The first letter refers to the chapter in which the code is contained and the subsequent two or three numbers refer to a related group of diseases or procedures and then specific disease or procedure within that group. For example, ICD-10 code K35.0 refers to acute appendicitis with generalized peritonitis. The K chapter is any disease of the digestive system and K35 group is all acute appendicitis. Cleary *et al.*²⁹ reported an accuracy of 51% at the four digit level but 90% at the three digit level suggesting that many inaccuracies occur at four digit level. For some uses, three digit accuracy (e.g. K35) may be sufficient. Three digit accuracy will be higher than described in this study.

What this study adds

The accuracy reported in this study is lower than previously reported³ and variable with a median of 90%. The current study contains a larger number of more recent studies. It is difficult to assess how applicable these figures are to general accuracy rates in the NHS or whether they reflect a degree of publication bias. Clinical studies that demonstrate good data accuracy may not be published with the aim of assessing data accuracy but focus on examining a particular clinical condition. Such articles may not be included in this analysis. Similarly, some articles that demonstrate poor data accuracy may have originally been conceived to look at a particular condition thereby skewing results towards a lower overall accuracy rate. The latest audit of data quality from the Audit Commission concluded that the accuracy of data coding was improving each year suggesting that there is discrepancy between published figures and real-life data accuracy.³⁰

If we accept the 87% overall accuracy reported by the Audit Commission, what are the possible uses of administrative data within the NHS? HES had been used for epidemiological and outcome-based research.^{43–48} It is difficult to quantify the impact of this accuracy level on research. An assumption is made that there are no systematic inaccuracies. A study, which examines the impact of an explanatory variable on outcome, assumes that the level of inaccuracy will be the same across that variable. This will be impossible to measure without a large NHS wide survey of all trust across

all specialities. Such a study would be expensive but may be possible through data collected by the Audit Commission National Audit. It is important that the current focus on improving data quality continues despite the proposed disbandment of the Audit Commission.

Several studies and the Audit Commission report examined the effect of data inaccuracy on reimbursement.^{20,30,31} Potential savings for individual trusts are considerable. One study estimated that inaccurate coding could lead to losses of up to 10% of department profits.³¹ It is in the interests of trusts to maximize their financial returns but important that data are as accurate as possible given the temptation to use codes associated with maximum financial return. Such ‘gaming’ should be avoided. In conjunction with outcome-based research, administrative data offer an attractive source for quality measurement. Poor quality data collection may reflect more widespread system failures within trusts or departments. Caution should be exercised regarding the reliability of identification of outliers from routinely collected data with outlier status serving as a prompt for further investigation rather than a definitive assertion of poor performance.

The introduction of PbR led to an improvement in diagnosis accuracy. Factors such as efficiency of hospital support systems, differences in unit case mix, organizational culture or management structure may further underlie persisting variation. Further work is required to assess the impact of these factors.

This review seeks to assess data accuracy in Great Britain but increasingly routinely collected and registry data are being used to draw international comparisons of performance.⁴⁹ It is essential that when using both administrative and clinical registry databases that intercounty variations are well understood. Databases may not be comprehensive or may only include patients treated at centres of excellence with an interest in data collection. Attempts should be made in each country to address the issues of data accuracy outlined in this study to ensure that data may be meaningfully used to explore national differences.

Clinical registry versus administrative databases

Clinical registries are purpose-built databases for prospective data collection. In contrast to the inclusive mandatory administrative data sets, clinical registries are mostly voluntary. They will not include all patients with a given condition nor will data entry be complete.⁵⁰ Two studies found HES and registry data to have largely comparable mortality with larger patient volumes recorded on HES.^{36,37} Four studies, however, found fewer cases recorded on the administrative database than in the clinical registries.^{32–34,38} The reasons

for this discrepancy are uncertain. It may represent poor coding on the HES data set but there was considerable variation in classifications used between the two data sets. For example, the definition of mortality and included procedures differed between the HES and CCAD data sets in the study by Westaby *et al.*³⁸ Though registries contain clinically meaningful data, they are more expensive and require enthusiastic clinicians to support data submission. Costs of maintaining HES data have been estimated at £1 per record with clinical registry data costing up to £60 per record.⁵¹ Though useful in discrete conditions or for specific treatments, registries may not reflect the full range of procedures performed even within a given specialty as clinicians may favour the entry of 'interesting' or complex cases over more straightforward cases.

Limitations of this study

The accuracy of routinely collected data is infrequently published. This review includes studies over an extended time period. The historical nature of the data limits contemporary applicability. Though our review was as broad as possible, some studies that have not referenced 'accuracy' in the title or abstract will not have been included in the study. It is difficult to quantify the impact of such bias on the results of this study.

The included studies are heterogeneous. They vary in methods used to assess accuracy, the diagnoses and procedures included and the personnel involved in assessing the data quality. Meta-analysis was, therefore, not possible. Indeed due to the small number of papers, limited statistical analysis was possible. Few studies looked at accuracy in recent years following the introduction of PbR and concerted efforts to improve data quality. The wide range of data accuracies reported may reflect considerable variation in practice across the NHS or differences in methodologies used in the included studies. Inter-coder reliability was rarely stated in these studies. Only 68% of the studies used random sampling and 48% of the studies stated that trained coders were used. Methods of identifying case records for review varied across studies. Some studies used local databases^{26,35} or all admissions in a defined period with or without a specific diagnosis or under a certain physician^{8,10,13–16,19–21,23–25,28–31,39} to identify included patients. Studies with accuracy rates at the extremes of the spectrum may be preferentially reported. Though given the wide range of accuracies reported, preference for low or high rates is likely to be limited. The overall accuracy reported in this study cannot be extrapolated to individual NHS trusts. Some trusts will have more reliable data than

others. Some diagnoses or procedures may be better coded than others. The clinician's diagnosis at discharge was the gold standard against which accuracy was measured. This relies on correct diagnosis at discharge. The diagnosis may be uncertain or become apparent later.

NHS administrative data accuracy has improved in recent years. This may relate to the introduction of *prorata* financial reimbursement. This review suggests that data accuracy is sufficient for use in most circumstances. Wide variation in reported accuracy may reflect variation in individual trusts' coding suggesting that care should be exercised when using these data for clinician and institution benchmarking. Identification of apparent unacceptable institution or individual performance using administrative data should serve as a prompt for further investigation and be interpreted with caution.

Funding

The Dr Foster Unit at Imperial is largely funded by a research grant from Dr Foster Intelligence (an independent health service research organization). The Unit is also affiliated with the Centre for Patient Safety and Service Quality at Imperial College Healthcare NHS Trust, which is funded by the National Institute of Health Research. We are grateful for support from the NIHR Biomedical Research Centre funding scheme.

References

- 1 World Health Organization. International Classification of Diseases (ICD). <http://www.who.int/classifications/icd/en/> (May 2011, date last accessed).
- 2 Connecting for Health. OPCS-4 Classification. http://www.connectingforhealth.nhs.uk/systemsandservices/data/clinicalcoding/codingstandards/opcs4/index_html (May 2011, date last accessed).
- 3 Campbell SE, Campbell MK, Grimshaw JM *et al.* A systematic review of discharge coding accuracy. *J Public Health Med* 2001;**23**(3):205–11.
- 4 Audit Commission. PbR data assurance framework 2007/08. <http://www.audit-commission.gov.uk/SiteCollectionDocuments/AuditCommissionReports/NationalStudies/PbRreport.pdf> (May 2011, date last accessed).
- 5 Department of Health. The operating framework for the NHS in England 2008/09. 2007. http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/@dh/@en/documents/digitalasset/dh_081271.pdf (May 2011, date last accessed).
- 6 Shea BJ, Grimshaw JM, Wells GA *et al.* Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC Med Res Methodol* 2007;**7**:10.

- 7 Crombie I. *The Pocket Guide to Critical Appraisal*. London: BMJ Publishing Group, 1996.
- 8 Colville RJ, Laing JH, Murison MS. Coding plastic surgery operations: an audit of performance using OPCS-4. *Br J Plast Surg* 2000;**53**(5):420–2.
- 9 Davenport RJ, Dennis MS, Warlow CP. The accuracy of Scottish Morbidity Record (SMR1) data for identifying hospitalised stroke patients. *Health Bull (Edinb)* 1996;**54**(5):402–5.
- 10 Dixon J, Sanderson C, Elliott P *et al*. Assessment of the reproducibility of clinical coding in routinely collected hospital activity data: a study in two hospitals. *J Public Health Med* 1998;**20**(1):63–9.
- 11 Dornan S, Murray FE, White G *et al*. An audit of the accuracy of upper gastrointestinal diagnoses in Scottish Morbidity Record 1 data in Tayside. *Health Bull (Edinb)* 1995;**53**(5):274–9.
- 12 Gibson N, Bridgman SA. A novel method for the assessment of the accuracy of diagnostic codes in general surgery. *Ann R Coll Surg Engl* 1998;**80**(4):293–6.
- 13 Harley K, Jones C. Quality of Scottish Morbidity Record (SMR) data. *Health Bull (Edinb)* 1996;**54**(5):410–7.
- 14 Hasan M, Meara RJ, Bhowmick BK. The quality of diagnostic coding in cerebrovascular disease. *Int J Qual Health Care* 1995;**7**(4):407–10.
- 15 Kirkman MA, Mahattanakul W, Gregson BA *et al*. The accuracy of hospital discharge coding for hemorrhagic stroke. *Acta Neurol Belg* 2009;**109**(2):114–9.
- 16 Kohli HS, Knill-Jones RP. How accurate are SMR1 (Scottish Morbidity Record 1) data? *Health Bull (Edinb)* 1992;**50**(1):14–23; discussion 29–31.
- 17 McGonigal G, McQuade C, Thomas B. Accuracy and completeness of Scottish mental hospital in-patient data. *Health Bull (Edinb)* 1992;**50**(4):309–14.
- 18 Mitra I, Malik T, Homer JJ *et al*. Audit of clinical coding of major head and neck operations. *Ann R Coll Surg Engl* 2009;**91**(3):245–8.
- 19 Murchison J, Barton JR, Ferguson A. An analysis of cases incorrectly coded as inflammatory bowel disease in Scottish Hospital In-Patient Statistics (SHIPS). *Scott Med J* 1991;**36**(5):136–8.
- 20 Nouraci SA, O'Hanlon S, Butler CR *et al*. A multidisciplinary audit of clinical coding accuracy in otolaryngology: financial, managerial and clinical governance considerations under payment-by-results. *Clin Otolaryngol* 2009;**34**(1):43–51.
- 21 Panayiotou B. Coding of clinical diagnoses. Persevere with Korner system. *BMJ* 1993;**306**(6891):1541.
- 22 Park RH, McCabe P, Russell RI. Who should log SHIPS? The accuracy of Scottish Hospital Morbidity Data for Wilson's disease. *Health Bull (Edinb)* 1992;**50**(1):24–8; discussion 29–31.
- 23 Pears J, Alexander V, Alexander GF *et al*. Audit of the quality of hospital discharge data. *Health Bull (Edinb)* 1992;**50**(5):356–61.
- 24 Reddy-Kolanu GR, Hogg RP. Accuracy of clinical coding in ENT day surgery. *Clin Otolaryngol* 2009;**34**(4):405–7.
- 25 Samy AK, Whyte B, MacBain G. Abdominal aortic aneurysm in Scotland. *Br J Surg* 1994;**81**(8):1104–6.
- 26 Sellar C, Goldacre MJ, Hawton K. Reliability of routine hospital data on poisoning as measures of deliberate self poisoning in adolescents. *J Epidemiol Community Health* 1990;**44**(4):313–5.
- 27 Smith SH, Kershaw C, Thomas IH *et al*. PIS and DRGs: coding inaccuracies and their consequences for resource management. *J Public Health Med* 1991;**13**(1):40–1.
- 28 Yeoh C, Davies H. Clinical coding: completeness and accuracy when doctors take it on. *BMJ* 1993;**306**(6883):972.
- 29 Cleary R, Beard R, Coles J *et al*. Comparative hospital databases: value for management and quality. *Qual Health Care* 1994;**3**(1):3–10.
- 30 Audit Commission. Improving data quality in the NHS Annual report on the PbR assurance programme Health 2010. <http://www.audit-commission.gov.uk/SiteCollectionDocuments/Downloads/26082010pbrnhsdataqualityreport.pdf> (May 2011, date last accessed).
- 31 Beckley IC, Nouraci R, Carter SS. Payment by results: financial implications of clinical coding errors in urology. *BJU Int* 2009;**104**(8):1043–6.
- 32 Jen MH, Holmes AH, Bottle A *et al*. Descriptive study of selected healthcare-associated infections using national Hospital Episode Statistics data 1996–2006 and comparison with mandatory reporting systems. *J Hosp Infect* 2008;**70**(4):321–7.
- 33 Mukherjee AK, Leck I, Langley FA *et al*. The completeness and accuracy of health authority and cancer registry records according to a study of ovarian neoplasms. *Public Health* 1991;**105**(1):69–78.
- 34 Raza Z, Holdsworth RJ, McCollum PT. Accuracy of the recording of operative events by the Scottish Morbidity Record 1 (SMR1) for a teaching hospital vascular unit. *J R Coll Surg Edinb* 1999;**44**(2):96–8.
- 35 Milburn JA, Driver CP, Youngson GG *et al*. The accuracy of clinical data: a comparison between central and local data collection. *Surgeon* 2007;**5**(5):275–8.
- 36 Garout M, Tilney HS, Tekkis PP *et al*. Comparison of administrative data with the Association of Coloproctology of Great Britain and Ireland (ACPGBI) colorectal cancer database. *Int J Colorectal Dis* 2008;**23**(2):155–63.
- 37 Aylin P, Lees T, Baker S *et al*. Descriptive study comparing routine hospital administrative data with the Vascular Society of Great Britain and Ireland's National Vascular Database. *Eur J Vasc Endovasc Surg* 2007;**33**(4):461–5; discussion 66.
- 38 Westaby S, Archer N, Manning N *et al*. Comparison of hospital episode statistics and central cardiac audit database in public reporting of congenital heart surgery mortality. *BMJ* 2007;**335**(7623):759.
- 39 Drennan Y. Data quality, patient classification systems, and audit: a recent study. In: *Current Perspectives in Healthcare Computing 1994*. Harrogate: BJHC Ltd, 1994:54–60.
- 40 Audit Commission. Data remember: improving the quality of patient-based information in the NHS 2002. <http://www.audit-commission.gov.uk/SiteCollectionDocuments/AuditCommissionReports/NationalStudies/datarember.pdf> (May 2011, date last accessed).
- 41 Care Quality Commission. <http://www.cqc.org.uk/> (May 2011, date last accessed).
- 42 Williams JG, Mann RY. Hospital episode statistics: time for clinicians to get involved? *Clin Med Res* 2002;**2**:34–7.

- 43 Bottle A, Aylin P. Application of AHRQ patient safety indicators to English hospital data. *Qual Saf Health Care* 2009;**18**(4):303–8.
- 44 Faiz O, Aylin P, Bottle A. Changing trends in surgery for acute appendicitis (Br J Surg 2008; 95: 363–368). *Br J Surg* 2008;**95**(6):801; author reply 01.
- 45 Faiz O, Blackburn SC, Clark J *et al.* Laparoscopic and conventional appendectomy in children: outcomes in English hospitals between 1996 and 2006. *Pediatr Surg Int* 2008;**24**(11):1223–7.
- 46 Faiz O, Brown T, Bottle A *et al.* Impact of hospital institutional volume on postoperative mortality after major emergency colorectal surgery in English National Health Service Trusts, 2001 to 2005. *Dis Colon Rectum* 2010;**53**(4):393–401.
- 47 Faiz O, Haji A, Burns E *et al.* Hospital stay amongst patients undergoing major elective colorectal surgery: predicting prolonged stay and readmissions in NHS hospitals. *Colorectal Dis* 2011;**13**(7): 816–22.
- 48 Mayer EK, Bottle A, Darzi AW *et al.* The volume-mortality relation for radical cystectomy in England: retrospective analysis of hospital episode statistics. *BMJ* 2010;**340**:c1128.
- 49 Coleman MP, Forman D, Bryant H *et al.* Cancer survival in Australia, Canada, Denmark, Norway, Sweden, and the UK, 1995-2007 (the International Cancer Benchmarking Partnership): an analysis of population-based cancer registry data. *Lancet* 2011;**377**(9760):127–38.
- 50 National Bowel Cancer Audit 2009. <http://www.ic.nhs.uk/webfiles/Services/NCASP/audits%20and%20reports/National%20Bowel%20Cancer%20Exec%20Summary%202009%2005-11-09.pdf> (May 2011, date last accessed).
- 51 Raftery J, Roderick P, Stevens A. Potential use of routine databases in health technology assessment. *Health Technol Assess* 2005;**9**(20): 1–466. <http://www.hta.ac.uk/fullmono/mon920.pdf> (May 2011, date last accessed).