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, W2 1NY 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
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. 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.

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.

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 Crombie and utilized previously by Campbell et al. 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 training or experience was specifically mentioned; ‘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.
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 and seven studies contrasted routinely collected data with clinical registry data. 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 and seven studies contrasted routinely collected data with clinical registry data. The clinicians’ diagnosis at discharge was the standard against which accuracy was measured.

![Table 1](https://academic.oup.com/jpubhealth/article-abstract/34/1/138/1553464/140-JOURNAL-OF-PUBLIC-HEALTH)
diagnostic coding \(^8\), \(^{12,14\text{-}17,19\text{-}23,25,26,28,29}\) 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\textendash}92.1\%). The median diagnostic accuracy was 80.3\% (IQR: 63.3\textendash}94.1\%) with a median procedure accuracy of 84.2\% (IQR: 68.7\textendash}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 \([\text{pre-PbR} 77.0\% \text{ (IQR: 66.2\textendash}89.0\% \text{ versus post-PbR 86.1\% \text{ (IQR: 73.1\textendash}96.1\%)}, P = 0.207\]) or the accuracy of procedure codes \((P = 0.602)\) but the accuracy of the primary diagnosis improved \([73.8\% \text{ (IQR: 59.3\textendash}92.1\%) \text{ versus 96.0\% \text{ (IQR: 89.3\textendash}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 \((P = 0.292, P = 0.245 \text{ and } P = 0.742, \text{ respectively})\).

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

**Papers comparing routinely collected data with clinical registry data**

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

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 \((\text{HES } n = 16,923 \text{ and NVD } n = 8,462)\) with slightly higher death rates recorded on HES \((\text{HES}, 18\% \text{ and NVD, 15\%})\).\(^{37}\) 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>
<thead>
<tr>
<th>Country</th>
<th>Study</th>
<th>Year</th>
<th>Diagnosis/procedure included</th>
<th>Study dates</th>
<th>Coding system</th>
<th>Number of cases sampled</th>
<th>Setting</th>
<th>Data accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>England</td>
<td>Smith et al.</td>
<td>1991</td>
<td>Joint replacements</td>
<td>1988</td>
<td>OPCS3</td>
<td>139</td>
<td>3 hospitals</td>
<td>Procedure, 85.0%</td>
</tr>
<tr>
<td>England</td>
<td>Panayiotou</td>
<td>1993</td>
<td>Cerebrovascular disease</td>
<td>Unspecified</td>
<td>ICD-9</td>
<td>117</td>
<td>Single acute hospital</td>
<td>Diagnosis, 76.0%</td>
</tr>
<tr>
<td>England</td>
<td>Cleary et al.</td>
<td>1994</td>
<td>All general medicine and general surgery diagnoses</td>
<td>1990–1991</td>
<td>OPCS4.2</td>
<td>2044</td>
<td>4 acute hospitals</td>
<td>Diagnosis, 68.0%</td>
</tr>
<tr>
<td>England</td>
<td>Drennan</td>
<td>1994</td>
<td>Urology, cardiothoracics, cardiology, general surgery diagnoses</td>
<td>1990–1991</td>
<td>OPCS4, ICD-9</td>
<td>298</td>
<td>Single acute hospital</td>
<td>Diagnosis, 71.0%</td>
</tr>
<tr>
<td>England</td>
<td>Gibson and Bridgman</td>
<td>1998</td>
<td>General surgery diagnosis</td>
<td>1995</td>
<td>ICD-10</td>
<td>298</td>
<td>Diagnosis, 1252; procedure, 416</td>
<td>Diagnosis, 50.5%</td>
</tr>
<tr>
<td>England</td>
<td>Kirkman et al.</td>
<td>2009</td>
<td>Haemorrhagic stroke</td>
<td>2002–2007</td>
<td>ICD-10</td>
<td>298</td>
<td>ICH 978, SAH 1169</td>
<td>Single acute hospital</td>
</tr>
<tr>
<td>England</td>
<td>Reddy-Kolanu and Hogg</td>
<td>2009</td>
<td>ENT procedures</td>
<td>2008</td>
<td>OPCS4</td>
<td>79</td>
<td>Hospital day surgery unit</td>
<td>Procedure, 69.6%</td>
</tr>
<tr>
<td>England</td>
<td>Mitra et al.</td>
<td>2009</td>
<td>Head and neck surgery procedures</td>
<td>2006</td>
<td>OPCS</td>
<td>34</td>
<td>Single acute hospital</td>
<td>Procedure, 85.1%</td>
</tr>
<tr>
<td>England</td>
<td>Beckley et al.</td>
<td>2009</td>
<td>Urological procedures</td>
<td>2007</td>
<td>ICD-10, OPCS4</td>
<td>500</td>
<td>Single acute hospital</td>
<td>Procedure, 90.9%</td>
</tr>
<tr>
<td>England</td>
<td>Audit Commission</td>
<td>2010</td>
<td>All diagnoses</td>
<td>2009–2010</td>
<td>ICD-10, OPCS4</td>
<td>Unknown</td>
<td>Single acute hospital</td>
<td>Procedure, 83.4% Overall, 87.0% Diagnosis, 87.0% Procedure, 90.0% Overall, 93.7% Crohn’s, 95.5% Ulcerative colitis, 91.0%</td>
</tr>
<tr>
<td>Scotland</td>
<td>Murchison et al.</td>
<td>1991</td>
<td>Inflammatory bowel disease</td>
<td>1968–1983</td>
<td>ICD-8, ICD-9</td>
<td>255</td>
<td>All NHS hospitals in Scotland</td>
<td>Diagnosis, 87.50%</td>
</tr>
<tr>
<td>Scotland</td>
<td>Park et al.</td>
<td>1992</td>
<td>Wilson’s disease</td>
<td>1974–1989</td>
<td>ICD-8, ICD-9</td>
<td>40</td>
<td>All Scotland</td>
<td>Diagnosis, 73.6%</td>
</tr>
<tr>
<td>Scotland</td>
<td>Kohli et al.</td>
<td>1992</td>
<td>Gastrointestinal Diagnosis, co-existing Arthritis</td>
<td>1987</td>
<td>ICD-9</td>
<td>778</td>
<td>Multiple hospitals</td>
<td>Diagnosis, 93%</td>
</tr>
<tr>
<td>Scotland</td>
<td>Pears et al.</td>
<td>1992</td>
<td>Paediatric and general medical diagnoses</td>
<td>Unspecified</td>
<td>ICD-9</td>
<td>100</td>
<td>Single hospital</td>
<td>Medical diagnosis, 54.0%</td>
</tr>
</tbody>
</table>
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].

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. Steps have been taken to improve quality. The Care Quality Commission mandates yearly audits of individual trust data quality. 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 report, 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. Clinician involvement in coding has been proposed to improve accuracy. Yeoh and Davies 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. 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

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

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.

*Definition of 30-day mortality and breadth of included procedures varied between HES and CCAD.*
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. 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 reported 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.

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. 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. Potential savings for individual trusts are considerable. One study estimated that inaccurate coding could lead to losses of up to 10% of department profits. 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. 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. Two studies found HES and registry data to have largely comparable mortality with larger patient volumes recorded on HES. Four studies, however, found fewer cases recorded on the administrative database than in the clinical registries.
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 databases26,35 or all admissions in a defined period with or without a specific diagnosis or under a certain physician8,10,13–16,19–21,23–25,28–31,39 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


