Monitoring the use and outcomes of new devices and procedures: how does coding affect what Hospital Episode Statistics contribute? Lessons from 12 emerging procedures 2006–10

Hannah Patrick1, Andrew Sims2, Julie Burn2, Derek Bousfield2, Elaine Colechin2, Christopher Reay2, Neil Alderson2, Stephen Goode3, David Cunningham4, Bruce Campbell1,5

1National Institute for Health and Clinical Excellence, London, UK
2Newcastle upon Tyne Hospitals NHS Foundation Trust, Newcastle, UK
3Sheffield Vascular Institute, Northern General Hospital, Sheffield, UK
4National Institute for Cardiovascular Outcomes Research, London, UK
5National Institute for Health and Clinical Excellence, and Peninsula College of Medicine & Dentistry, Exeter and Plymouth, UK

Address correspondence to H. Patrick, E-mail: hannah.patrick@nice.org.uk

ABSTRACT

Background New devices and procedures are often introduced into health services when the evidence base for their efficacy and safety is limited. The authors sought to assess the availability and accuracy of routinely collected Hospital Episodes Statistics (HES) data in the UK and their potential contribution to the monitoring of new procedures.

Methods Four years of HES data (April 2006–March 2010) were analysed to identify episodes of hospital care involving a sample of 12 new interventional procedures. HES data were cross checked against other relevant sources including national or local registers and manufacturers’ information.

Results HES records were available for all 12 procedures during the entire study period. Comparative data sources were available from national (5), local (2) and manufacturer (2) registers. Factors found to affect comparisons were miscoding, alternative coding and inconsistent use of subsidiary codes. The analysis of provider coverage showed that HES is sensitive at detecting centres which carry out procedures, but specificity is poor in some cases.

Conclusions Routinely collected HES data have the potential to support quality improvements and evidence-based commissioning of devices and procedures in health services but achievement of this potential depends upon the accurate coding of procedures.

Keywords health intelligence, health services, quality, health technology adoption, outcome assessment

Introduction

Although randomized controlled trials of new medical devices and interventional procedures are possible,1 testing under formal and rigid trial conditions is difficult because outcomes depend not only on the technology itself, but also on the skill of the operator and the setting in which the intervention is performed.2 In addition, in comparison with pharmaceuticals, there is less regulatory demand for clinical trials of devices and interventional procedures.2 As a result
new devices and procedures are often introduced into medical practice when the evidence base for their efficacy, safety and cost-effectiveness is more limited in quantity and quality than for pharmaceuticals. Ongoing monitoring of the use of, and outcomes from, new interventions is particularly important in this setting of limited evidence base. Several publications have highlighted the importance of monitoring outcomes from new interventions and improving the process by which new technologies and techniques enter into practice.3,4

In the UK, the National Institute for Health and Clinical Excellence (NICE) established its Interventional Procedures (IP) Programme in 2002 to support clinicians, healthcare organizations and the NHS in the process of introducing new procedures. If the evidence is inadequate then a recommendation is usually made for use only with ‘special arrangements’ for clinical governance, consent and audit. ‘Special arrangements’ guidance has been published following 168 of the first 368 evaluations. While this stipulates ongoing audit of outcomes, it has been recognized from the outset that the ideal would be collection of data on each procedure done, including both safety and efficacy outcomes, when the evidence is inadequate.

Safety and efficacy data for new procedures have several possible uses, in addition to providing the kind of evidence which contributes to evaluation by NICE:

- Management of the risk is associated with the introduction of new procedures, including the monitoring of outcomes for new operators on the learning curve.
- Health service quality monitoring and demonstration of quality standards5
- Use by individual clinicians to inform clinical audit and demonstrate continuing professional competency.
- Informing Health Service Commissioning decisions (with the ultimate aim of evaluating how resources used relate to services delivered and health improvements achieved).

To achieve these objectives it is necessary to collect national as well as local data, and national registers have often been proposed as the method to achieve this.6,7 But despite many references in the literature to their usefulness, registers have been poorly resourced and routine monitoring is rarely done other than as part of post-market surveillance by manufacturers. Reasons for this dearth of action may include the lack of resources to enable the data collection and submission, and scepticism about the quality of data.8 As a result, many registers reflect only the data of a keen minority of clinicians who have managed to secure sufficient resources to assist in collating data for submission.9

An alternative strategy is optimization of the use of existing routine health services data. In the UK, routine healthcare data are collected in the Hospital Episode Statistics (HES).10 They use the Office of Population Censuses and Survey (OPCS-4) Classification of Surgical Operations which is supported, maintained and developed by the NHS Classification Service (NCS). This service has worked increasingly with NICE to recommend appropriate OPCS-4 codes for procedures covered by NICE guidance and to develop new codes where none already exist.11,12

The aims of this study were to assess the availability and accuracy of routinely available HES data as a tool to monitor the introduction of new interventional procedures into practice and to investigate whether the coverage of the data for individual procedures is affected by the complexity and specificity of their OPCS-4 codes.

**Methods**

Twelve interventional procedures were selected: 11 from published NICE Interventional Procedure Guidance (IPG) and one without NICE guidance (iliac artery stenting) but suggested by a professional society. The procedures were chosen to cover a parameter space defined by three dimensions: approximate number of procedures carried out per year, number of hospitals in which they were likely to be done and complexity of their OPCS-4 coding (some procedures have more than one possible choice of code, some require an additional code to specify the site of the operation and some have non-specific coding that does not allow the identification of the exact procedure separately from other similar procedures). Table 1 shows the

**Table 1** Distribution of the 12 procedures selected for analysis

<table>
<thead>
<tr>
<th></th>
<th>Number in sample of 12 procedures</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Number of procedures per year</strong></td>
<td></td>
</tr>
<tr>
<td>&lt;100</td>
<td>4</td>
</tr>
<tr>
<td>100 to &lt;500</td>
<td>5</td>
</tr>
<tr>
<td>500+</td>
<td>3</td>
</tr>
<tr>
<td><strong>Number of hospitals</strong></td>
<td></td>
</tr>
<tr>
<td>&lt;10</td>
<td>2</td>
</tr>
<tr>
<td>10 to &lt;50</td>
<td>7</td>
</tr>
<tr>
<td>50+</td>
<td>3</td>
</tr>
<tr>
<td><strong>Complexity of coding</strong></td>
<td></td>
</tr>
<tr>
<td>Single code, specific</td>
<td>4</td>
</tr>
<tr>
<td>Multiple or non-specific</td>
<td>6</td>
</tr>
<tr>
<td>Site code needed</td>
<td>2</td>
</tr>
</tbody>
</table>
distributions of the 12 selected procedures across the three dimensions.

Recommended OPCS and International Classification of Diseases (ICD) coding for each procedure was taken from the NICE website wherever possible. For the few cases where the website did not provide coding recommendations, codes were taken from the OPCS-4 tabular list and consensus achieved with the study group as to the coding that would be used to interrogate HES. In some cases national registers use coding other than that recommended on the NICE website. Alternatives were recorded and used in the HES analysis in order to test the effect of coding methodology on HES coverage. The OPCS-4 procedure codes used to identify the 12 procedures in HES are shown in Table 2.

The existence of any register for any of the 12 procedures was investigated by contacting relevant clinicians and their professional organizations and by doing web searches. For some procedures there was no national register but clinicians offered data from local sources: in these cases, the subset of the HES data for the relevant hospital/s was used to inform comparisons between different data sources.

Where we failed to identify any national or local data set, but a procedure required a specific device(s), relevant manufacturers were contacted to ask for sales data. Manufacturers were contacted by telephone, letters or e-mails and asked to provide UK sales figures broken down by financial year (2006–10) and by hospital. Confidentiality agreements were drawn up as required.

Data analysis
HES data were extracted for all 12 procedures, for 4 financial years (2006–10) based on year of finished consultant episode (defined as the period of time a patient spends in hospital under the care and responsibility of one consultant team) and were imported into a local, securely held, Structured Query Language database for analysis. Data extraction queries from HES were written using the data fields selected from the online HES Data Dictionaries.

Each episode of care recorded on HES includes data fields that enable the capture of coding for multiple interventional procedures. For the purposes of this study, HES data for 10 levels of procedural coding were interrogated and if a relevant OPCS code was found in any one of these fields then that episode of care was included in the comparative analysis with other data sources. Comparative analyses were restricted to healthcare facilities in England.

National registers aim to achieve comprehensive coverage but they do not provide a ‘gold standard data set’ and therefore the sensitivity of data was analysed (i) using register data as the reference data set and then (ii) using HES data as the reference data set.

Quality assurance
As a check of data quality, prior to undertaking any detailed analysis, the quantity of relevant episodes of care in the HES extract was checked at an aggregate level against data available from the HESonline website, which provides summary figures for each year, including numbers of finished consultant episodes broken down by procedure code.

In addition, for register data, two of the authors independently calculated the total numbers of procedures.

Results
Comparative data sources were available for 9 of the 12 procedures from national (5), local (2) and manufacturer (2) registers. Three out of the four procedures for which NICE recommended ‘Special arrangements’ for data collection had no organized data collection at a national level. Summary HES data for all 12 procedures for all 4 financial years are presented in Table 2.

For the comparisons between HES and national registers (five procedures), the number of hospitals found to be undertaking the procedure through HES always exceeded the number reporting to the register. For example HES identified 33 hospitals undertaking Endovascular closure of atrial septal defect (IPG96) but 26 hospitals submitted data to the national Central Cardiac Audit Database (CCAD).

Like-for-like comparisons of data from matched sources by year and by hospital were possible for nine procedures (Table 2). This analysis demonstrated the difficulties of data interrogation when using a complex, hierarchical coding system.

OPCS coding rules require that some procedures are identified by a combination of procedural code and anatomical site (a Z code). For example, the OPCS code recommended for NICE guidance IPG67 (balloon dilatation of pulmonary valve stenosis) is K35.5 (‘percutaneous transluminal valvuloplasty’) combined with Z32.4 (‘pulmonary valve’). Table 2 presents an analysis of HES data showing that this combination of codes yielded just 220 relevant episodes of care. The national register (the CCAD) records the procedure by OPCS code K35.4 (‘percutaneous transluminal pulmonary valvotomy including percutaneous transluminal balloon valvotomy of pulmonary valve’). We repeated the HES analysis using K35.4, and after adjusting for episodes
Table 2 Summary of procedures, recommendations, data availability and matched comparisons

<table>
<thead>
<tr>
<th>NICE guidance†</th>
<th>Procedure description and OPCS-4 code</th>
<th>Comparator availability</th>
<th>Centres</th>
<th>Hospitals</th>
<th>Episodes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>HES</td>
<td>Comparator</td>
<td>HES</td>
</tr>
<tr>
<td>IPG329, SA</td>
<td>Total prosthetic replacement of the temporomandibular joint, V20.1</td>
<td>Data available independently from two hospitals (one centre provided data for 2009/10 only)</td>
<td>19</td>
<td>2</td>
<td>122</td>
</tr>
<tr>
<td>IPG79, N</td>
<td>Stent placement for vena caval obstruction, L79.3</td>
<td>Local data provided by two hospitals (one centre provided data for 2007/08–2009/10 only)</td>
<td>125</td>
<td>2</td>
<td>1365</td>
</tr>
<tr>
<td>IPG97, N</td>
<td>Endovascular closure of patent ductus arteriosus, L03.1</td>
<td>Register data from CCAD with relevant episodes identified by CCAD database manager</td>
<td>31</td>
<td>16</td>
<td>1655</td>
</tr>
<tr>
<td>IPG96, N</td>
<td>Endovascular closure of atrial septal defect, K13.3</td>
<td>Register data from CCAD with relevant episodes identified by CCAD database manager</td>
<td>33</td>
<td>26</td>
<td>1702</td>
</tr>
<tr>
<td>IPG237, SA</td>
<td>Percutaneous pulmonary valve implantation, K35.7</td>
<td>Register data from CCAD with relevant episodes identified by CCAD database manager</td>
<td>11</td>
<td>6</td>
<td>118</td>
</tr>
<tr>
<td>IPG67, N</td>
<td>Balloon dilatation of pulmonary valve stenosis, K35.4</td>
<td>Register data from CCAD with relevant episodes identified by CCAD database manager</td>
<td>15</td>
<td>13</td>
<td>477</td>
</tr>
<tr>
<td></td>
<td>K35.5 with Z32.4</td>
<td></td>
<td>17</td>
<td>13</td>
<td>220</td>
</tr>
<tr>
<td></td>
<td>Total (K35.4 + K35.5 with Z32.4)</td>
<td></td>
<td>13</td>
<td>215</td>
<td>832</td>
</tr>
<tr>
<td></td>
<td>—</td>
<td></td>
<td>19</td>
<td>13</td>
<td>695</td>
</tr>
<tr>
<td></td>
<td>Percutaneous transluminal insertion of stent into iliac artery, L54.4</td>
<td>Register data from BIAS with relevant episodes identified by whether a stent was implanted</td>
<td>133</td>
<td>39</td>
<td>7700</td>
</tr>
<tr>
<td>IPG64, N</td>
<td>Sacral nerve stimulation for urge incontinence and urge-frequency, A70.1 with Z11.2</td>
<td>—</td>
<td>38</td>
<td>3556</td>
<td>1447</td>
</tr>
<tr>
<td>IPG188, N</td>
<td>Deep brain stimulation for tremor and dystonia (excluding Parkinson’s disease), A09.1</td>
<td>Sales figures for 22 centres from a single manufacturer covering 3 calendar years 2007, 2008 and 2009</td>
<td>16</td>
<td>14</td>
<td>266</td>
</tr>
<tr>
<td>IPG209, SA</td>
<td>Implantation of accommodating intraocular lenses for cataract, C75.1</td>
<td>Accommodating lens sales figures for 56 centres from one manufacturer (market share not known)</td>
<td>121</td>
<td>56</td>
<td>15 050</td>
</tr>
<tr>
<td>IPG232, SA</td>
<td>Serial transverse enteroplasty procedure (STEP) for bowel lengthening in parenteral nutrition-dependent children, G78.8 with Y26.3</td>
<td>International registry identified but data were not available for comparison</td>
<td>3</td>
<td>—</td>
<td>3</td>
</tr>
<tr>
<td>IPG351, N</td>
<td>Stapled transanal rectal resection for obstructed defaecation syndrome, H41.2 with Y26.3</td>
<td>Register ran from 2006 to 08, but no episode-level data available</td>
<td>28</td>
<td>Unknown</td>
<td>281</td>
</tr>
</tbody>
</table>

†NICE guidance recommendations: N, normal arrangements; SA, special arrangements.

‡Only HES episodes with a primary ICD code relating to the urinary system were included.

§Only HES episodes with a primary ICD code relating to tremor or dystonia were included.

¶Comparator episode numbers represent 39% of total sales (estimated percentage relating to tremor and dystonia).

‖Only non-NHS patients counted, as this procedure is not performed by the NHS, though it is performed in NHS hospitals.

¶Estimate.
of care in which both codes had been specified, HES identified 685 of the 832 procedures that had been reported to CCAD by 13 matched hospitals (each step of the analysis is presented in Table 2).

The specificity of data also presented problems when comparing manufacturers’ sales figures with HES data, for example with regard to IPG188, ‘deep brain stimulation for tremor and dystonia excluding Parkinson’s disease’ (OPCS code A09.1). HES records included 1051 episodes of care relating to A09.1 and analysis of the primary diagnosis ICD codes showed that 266 episodes were for tremor and dystonia alone (rather than Parkinson’s disease). Between 2007 and 09 there was only one manufacturer supplying devices for use in this procedure but sales figures for this period did not differentiate between devices used to treat Parkinson’s disease and those for tremor and dystonia alone. In order to compare HES and sales data we analysed all HES episodes for A09.1, for calendar years 2007–09, and found that 39% of episodes were for treatment of tremor and dystonia (not associated with Parkinson’s disease). We applied this factor to the sales figures to estimate the number of procedures associated with the scope of IPG188 and present the data in Table 2.

Sensitivity analysis of coverage is presented in Table 3. This shows that HES is sensitive at detecting hospitals which do procedures of interest but its specificity is poor. For each of the five procedures covered by national registers as well as HES, we found hospitals which reported only one episode of care involving procedures on HES in the 4-year study period. We consider these instances as likely to represent coding errors and Table 3 presents revised coverage sensitivities for registers after removal of hospitals with only one reported episode of care involving the procedure from the HES totals. This demonstrates that some registers have poor coverage even when hospitals recording presumed coding errors are removed.

For one procedure (IPG329: total prosthetic replacement of the temporomandibular joint) HES identified 19 hospitals undertaking the procedure, but clinicians involved with establishing a national register considered that only 10 of these were likely to be correctly coded cases. Analysis of diagnostic codes revealed five cases that were unlikely to be related to replacement of the temporomandibular joint, all recorded by hospitals unknown to the specialist register and likely to be due to human error. For example one case was associated with a primary diagnosis of congestive heart failure, requiring procedure U20.1 (transthoracic echocardiography)—very similar to code V20.1 (total prosthetic replacement of the temporomandibular joint).

### Discussion

Our findings demonstrate that for procedures with simple specific codes (i.e. not requiring complex combinations of codes to describe the procedure), HES can accurately identify hospitals using new procedures and the numbers of those procedures undertaken. In contrast, HES data show poor specificity for procedures requiring complex combinations of OPCS coding.

<table>
<thead>
<tr>
<th>Procedure and register</th>
<th>Coverage sensitivity, % (95% CI)</th>
<th>Centres reporting to HES in error (total centres)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HES versus Register</td>
<td>Register versus HES (all activity)</td>
</tr>
<tr>
<td>IPG97: Endovascular closure of patent ductus arteriosus and CCAD</td>
<td>93.8 (69.8–99.8)</td>
<td>48.4 (30.2–66.9)</td>
</tr>
<tr>
<td>IPG96: Endovascular closure of atrial septal defect and CCAD</td>
<td>96.2 (80.4–99.9)</td>
<td>75.8 (57.7–88.9)</td>
</tr>
<tr>
<td>IPG237: Percutaneous pulmonary valve implantation and CCAD</td>
<td>83.3 (35.9–99.6)</td>
<td>45.5 (16.7–76.6)</td>
</tr>
<tr>
<td>IPG67: Balloon dilatation of pulmonary valve stenosis and CCAD</td>
<td>100.0 (75.3–100.0)</td>
<td>68.4 (43.4–87.4)</td>
</tr>
<tr>
<td>LS4.4 Percutaneous transluminal insertion of stent into iliac artery and BIAS</td>
<td>97.4 (86.5–99.9)</td>
<td>28.6 (21.1–37.0)</td>
</tr>
</tbody>
</table>

*Using centres reporting to register as the reference source.
The comparison of HES data with local clinical data sets demonstrated that HES may help to identify hospitals that have not registered cases on national databases; some of these cases represent coding errors but many represent true under-coverage by national registers. Detailed comparison of data between HES and CCAD demonstrated that the OPCS-4 system includes alternative, legitimate coding of individual procedures: unless all possible coding strategies are included in an analysis, figures derived from HES may undercount the true incidence of a procedure.

Several other authors have explored the potential for HES to make a more significant contribution to health service evaluation than at present. However, ongoing scepticism amongst clinicians about the validity and accuracy of the routine data has been reported and it has been suggested that, on its own, HES contains insufficient clinical detail to be used to monitor health service outcomes. Nevertheless, Aylin et al. used HES to develop a clinical model for prediction of mortality that demonstrated similar discrimination to clinical databases for three common cardiovascular and cancer-related procedures. Using HES data, Faiz et al. demonstrated that laparoscopic appendicectomy was associated with lower postoperative mortality rates than open appendicectomy but higher subsequent readmission rates. While HES may offer this kind of useful information, concerns remain about the accuracy and coverage of HES data for specific interventions. Aylin et al. compared the volume and outcomes in vascular surgery between HES and a clinical database (National Vascular Database) for patients undergoing either repair of abdominal aortic aneurysm, carotid endarterectomy or infrainguinal bypass. They showed significant differences in total numbers between the databases. Agreement was greater for data comparisons relating to consultants known to be submitting data to the national clinical database. For a novel procedure (minimally invasive repair of pectus excavatum), Patrick et al. found that HES had lower specificity but higher sensitivity in identifying relevant procedures compared with a national register.

Despite their shortcomings, HES data may be useful for work aimed at improving the quality of national registers. For example, this has been successfully achieved in the National Bowel Cancer Audit Project by the Association of Coloproctology of Great Britain and Ireland, which used HES to check the coverage of the Audit, and the UK National Joint Register which demonstrated important variations in hip and knee replacement revision rates through linkage of its data to HES.

A major difficulty in designing our study was the lack of a gold standard against which to assess the true number of procedures done. It would seem reasonable to assume that databases kept or managed by clinicians might be of higher quality than routinely collected statistics such as HES. But coverage of clinical databases has been shown to be patchy and Colville et al. demonstrated that the quality of coding by clinicians in operating theatres was worse than that recorded by the hospital coders. A recent systematic review of discharge coding has shown that accuracy of coding has improved since the introduction of payment by results in 2002. By triangulating the various data sources available for this study we aimed to get the best indication of the true use of procedures in hospitals within England.

Another possible source of error when using HES data is inter-operator variability in the methodology used to capture data from HES. In order to assess whether our data were subject to this error we checked our aggregated data against those available on HESonline. Finally, our work focused on a small sample of procedures evaluated by the NICE IP Programme which has, historically, relied upon notification of new procedures to its web portal for the identification of new interventions with uncertain efficacy or safety. It is possible that other procedures are introduced into patient care without the formal coding and review mechanism that this Programme offers. Neither the NICE IP Programme nor the current coding systems have a remit to review or identify minor modifications to existing interventions. These present a considerable challenge to quality monitoring systems.

Conclusions

Routinely collected health service data have the potential to provide evidence about new devices and procedures, which would support evidence-based commissioning and quality improvement initiatives in health services. Achieving this potential requires improvements in the simplicity and specificity of coding procedures (particularly when they are being used for particular indications). A higher priority should be given to the development of a reactive coding system with improved specificity of its codes.

Acknowledgements

We thank Georgios Lyratzopoulos for advice and help in preparation of the manuscript, Mirella Marlow, Programme Director at NICE, for her support and advice during work associated with this article, manufacturers who have contributed data, Andrew Brooke and Nicola Ingram of the NCS for coding advice and Craig Ramsay for statistical advice. We would particularly like to thank those who allowed us to use local data sources for comparisons with HES data: O
Bennett and A Sidebottom (Nottingham University Hospitals) and S Worrall (Bradford Hospital).

Conflict of interest statement
None of the authors have conflicts of interest to report. All authors contributed to data analysis, writing and editing of the manuscript.

Funding
B.C., H.P., S.G. have not received any funding from external sources for the conduct of this study. A.S., J.B., D.B., E.C., C.R. and N.A. were employed by Newcastle upon Tyne Hospitals NHS Foundation Trust which hosted an External Assessment Centre funded by NICE.

References
16 www.hesonline.nhs.uk (December 2011, date last accessed).