Can hospital episode statistics support appraisal and revalidation? Randomised study of physician attitudes

Giles P Croft, John G Williams, Robin Y Mann, David Cohen and Ceri J Phillips

Giles P Croft[1]
MB ChB MRCS,
Clinical Research
Fellow, Royal
College of
Physicians'
Information
Laboratory, Centre
for Health
Information,
Research and
Evaluation
(CHIRAL)

John G Williams[1]
FRCP, Professor of
Health Services
Research

Robin Y Mann[2]
MB ChB MSc,
Development
Manager

David Cohen[3]
BCom MPhilm,
Professor of Health
Economics

Ceri J Phillips[1]
MSc PhD, Professor
of Health
Economics, School
of Health Science

[1]University of
Wales Swansea
[2]Health Informatics
Unit, Royal College
of Physicians
[3]University of
Glamorgan, Wales

ABSTRACT – **Hospital episode statistics were originally designed to monitor activity and allocate resources in the NHS. Recently their uses have widened to include analysis of individuals' activity, to inform appraisal and revalidation, and monitor performance. This study investigated physician attitudes to the validity and usefulness of these data for such purposes, and the effect of supporting individuals in data interpretation. A randomised study was conducted with consultant physicians in England, Wales and Scotland. The intervention group was supported by a clinician and an information analyst in obtaining and analysing their own data. The control group was unsupported. Attitudes to the data and confidence in their ability to reflect clinical practice were examined before and after the intervention. It was concluded that hospital episode statistics are not presently fit for monitoring the performance of individual physicians. A more comprehensive description of activity is required for these purposes. Improvements in the quality of existing data through clinical engagement at a local level, however, are possible.**

KEY WORDS: appraisal, data quality, hospital episode statistics, physician attitudes, revalidation, secondary uses of data

Introduction

Information about patient care plays a vital role in the management of health services. Hospital Episode Statistics database (HES, England) and the Patient Episode Database for Wales (PEDW) capture a range of administrative and clinical information about patients who are admitted to hospital as a day case or inpatient. Collected since the 1980s, the two organisations were originally designed for allocating resources and service planning.[1]

The uses of HES and PEDW have grown to include performance monitoring,[2] informing clinical indicators,[3] and supporting Payment by Results in England.[4] Recent proposals suggest using these data to measure the activity of individual clinicians for

purposes of appraisal and revalidation,[5,6] and commercial services are available to do this.[7,8] There are, however, ongoing concerns about the validity of centrally submitted hospital activity data, particularly from the clinical perspective,[9] and poor data quality has been highlighted as one of the key risks for the successful implementation of Payment by Results.[10]

The Audit Commission considers one of the main reasons for poor data quality to be a lack of understanding among front line staff regarding the reasons for, and the benefits of, the routine information being collected.[11] The Commission recommended that more effort to involve clinical staff in validating and using hospital episode statistics would improve data quality, but while clinicians can request activity held in their name directly from HES/PEDW, the numbers to have done so in the past are negligible.[12]

The Royal College of Physicians' Information Laboratory (*i*Lab) was established in 2004 as a secure, controlled environment in which hospital episode statistics (HES/PEDW) could be readily accessed, analysed and shared with individual consultants, while supported by experts with an understanding of the data and of the clinical context. The facility is situated at the School of Medicine, University of Wales Swansea, with an electronic link to the Royal College of Physicians (RCP). The aim is to promote data quality improvement through clinical engagement.

A randomised study of hospital episode statistics was undertaken to:

- investigate the effect of this expert support on consultant attitudes to the validity and usefulness of the data
- investigate consultants' confidence in the ability of the data to reflect their clinical practice.

Methods

Consultant physicians in NHS hospital practice in England and Wales were identified from the workforce database held by the RCP, and invited by letter to take part. Those who returned a completed baseline questionnaire were sent further information about the study. Those who still wished to take part were randomised. Consultants found to have less

than 80 finished consultant episodes (FCEs) held against their name over a two year period were excluded from the randomised study. The intervention group was invited to attend the *i*Lab facilities in London or Swansea; controls were asked to seek data to support their next appraisal using their normal resources.

Data from hospital episode statistics (HES/PEDW) were obtained for each individual in the intervention group and a standard set of queries was run according to the consultant's specialty. Consultants were provided with a brief explanation of the data, its origins, analysis and uses, by a clinical research fellow and an information analyst. This was followed by presentation and discussion of the individual's analysed data. Post-intervention attitudes and confidence were measured by a second questionnaire. The control group completed a similar questionnaire three to six months after randomisation.

A second control group was established to measure the impact of providing access to the data without face-to-face support: consultants in Scotland were invited to use eSCRIPS (an online tool available for basic analyses of their NHS episodes of activity) to gather information in support of the appraisal process. Attitudes and confidence were measured in the same way.

The questionnaires were designed to obtain information about current levels of local engagement, participants' attitudes towards the data's validity and usefulness, and their confidence in its ability to reflect their practice. Inter-group variability for questions in baseline and follow-up questionnaires was calculated using the Mann-Whitney test for ordinal data. Within-group comparison of before and after data was compared using the Wilcoxon Signed-Rank Test.

Ethical approval was obtained from the RCP Ethics Committee and consent was obtained from all participants before their data were analysed.

Results

Although prior to the study no individual physicians had requested their data directly from HES or PEDW, 19% of the 6,102 physicians invited to participate in this study wished to take part. Table 1 describes the demographic characteristics of invitees and participants. Aside from a greater number of female consultants being excluded because of insufficient data, there were no differences in age, sex ratio, or travel time to an *i*Lab facility indicating that the sample was representative. There was a wide geographical distribution of all those who returned the baseline questionnaire. All main medical specialties were represented, and were equally distributed between groups, the exception being 50% of Scottish consultants who returned follow-up questionnaires being registered as general physicians.

Attitudes towards the validity and usefulness of routinely collected activity data were obtained from 1,331 physicians in England, Wales and Scotland (Table 2). Key findings were that 70% had never or rarely seen the coding of inpatients or day cases for their own patients and a large majority (>80%) reported no communication with coding staff regarding data held in their own name. Consultants had very little confidence in centrally held data and its ability to reflect their clinical practice: 81% had either no confidence or were not particularly confident that it could do so when asked. Popular reasons given for these low levels were previous audits of its quality, design limitations and a lack of confidence in the coding process (Table 3). Conversely many of those who expressed confidence in the data reported that they were already engaged in the processes of collection and validation.

Following the intervention, compared with those who received no active support (ie control and Scottish groups), physicians using the *i*Lab reported significant increases in:

Table 1. Demographic characteristics of consultant physicians.

	Number	M:F ratio	Mean age (standard error)	Mean travel time to *i*Lab facility
England and Wales				
All physicians on RCP database	6,352	3.8:1	48.9 (0.10)	1h 58 min
– Invited to participate	6,102	3.9:1	48.9 (0.10)	1h 58 min
– Wishing to participate	1,160	3.8:1	48.3 (0.20)	2h 07 min
– Excluded (FCEs <80)	225	1.8:1	48.5 (0.48)	NA
– Randomised	250	5.6:1	48.0 (0.41)	2h 13 min
– Completed intervention questionnaires	47	4.9:1	48.0 (1.02)	2h 13 min
– Completed control questionnaires	61	4.1:1	46.0 (0.76)	NA
Scotland				
– Invited to participate	653	4.6:1	48.8 (0.31)	NA
– Wishing to participate and included	135	3.5:1	48.2 (0.65)	NA
– Completed questionnaires	34	4.7:1	47.7 (1.48)	NA

F = female; FCEs = finished consultant episodes; *i*Lab = Information Laboratory; M = male; NA = not applicable; RCP = Royal College of Physicians.

- appreciation for the usefulness of the data
- awareness of the data collection process
- willingness to contribute to local data collection and validation
- intent to contact local information departments about the data held in their name
- likelihood of monitoring data quality
- likelihood of changing their practice concerning the collection/validation of data in the future (see Tables 4a and 4b for p values).

Those supported by the *i*Lab reported a significant increase in confidence in the data which was not seen in the two unsupported groups. Table 5 gives a summary of data quality issues described by consultants who attended the *i*Lab, some of which are discussed below.

Discussion

The *i*Lab project was established in order to understand the barriers and solutions to improving the use and quality of routinely collected activity data in hospitals. It has confirmed considerable

Table 2. Percentage responses to baseline questions from England, Wales and Scotland.

Question	Response	England % (n=1,128)	Wales % (n=63)	Scotland % (n=140)
How frequently do you see the activity data held in your name?	Never	17	22	21
	Rarely (less than once a year)	22	17	28
	Occasionally (once or twice a year)	32	25	39
	Quite regularly (monthly to every three months)	21	25	11
	Often (weekly to monthly)	5	6	1
	Not stated	2	3	1
How frequently do you see the coding of inpatient and/or day case episodes for patients?	Never	46	48	44
	Rarely (less than once a year)	26	21	26
	Occasionally (once or twice a year)	13	14	23
	Quite regularly (monthly to every three months)	8	10	5
	Often (weekly to monthly)	4	6	1
	Not stated	3	2	1
How frequently have you been approached by a member of the coding/information department concerning activity data held in your name?	Never	59	63	66
	Rarely (less than once a year)	22	17	23
	Occasionally (once or twice a year)	12	13	8
	Quite regularly (monthly to every three months)	4	5	2
	Often (weekly to monthly)	1	0	0
	Not stated	2	2	1
How frequently have you had occasion to approach the coding/information department concerning activity data held in your name?	Never	40	37	45
	Rarely (less than once a year)	27	30	31
	Occasionally (once or twice a year)	25	27	19
	Quite regularly (monthly to every three months)	5	5	3
	Often (weekly to monthly)	1	0	1
	Not stated	2	2	1
– How often does this result in a change in the coding of episodes?	Never	40	41	46
	Rarely	18	17	24
	Occasionally	12	10	6
	Often	6	5	2
	Always	2	3	2
	Not stated	22	24	19
How confident are you that data routinely coded for central returns accurately reflects your clinical activity?	No opinion	3	5	4
	No confidence	3	27	43
	Not particularly confident	42	52	42
	Reasonably confident	11	11	9
	Confident	2	3	2
	Very confident	2	0	0
	Not stated	3	2	1

n = number.

reservations about the data by physicians, a willingness to get involved by many, and potential benefits from such engagement.

Clinical confidence and engagement in data quality

The baseline results support the concept of a vicious circle of poor data quality: central returns data are perceived as being inaccurate and unfit for purpose by clinicians who remain disengaged from the use and validation of these data, and so their accuracy remains unchanged through neglect. Common reasons for this lack of confidence include poor communication between clinical and information staff, perceived inaccuracies in the data, and concerns regarding the dataset itself.

In Scotland a web-based service providing consultant-level activity data based on central returns is available to all. However, 86% of baseline respondents stated they were not particularly confident or had no confidence that these data reflected their clinical activity. If confidence in the data is higher in consultants who feel engaged with the process, simply providing data to consultants without active support cannot be considered as engagement. By contrast, the positive results from the intervention group suggest that engagement through active support is effective in improving confidence.

Local processes affecting the validity of national comparisons

Detailed examination of a consultant's data when compared with others highlighted a range of data quality issues and elicited strong views on its suitability for benchmarking purposes. Within each locality (ie hospital) marked differences not only in case mix, but also in admission, discharge and transfer policies, volume and type of staffing, administrative and clinical coding procedures, and the unique facilities provided by a hospital were found. These differences were reflected in the hospital statistics. While comparisons of activity between physicians *within* each locality were viewed as having potential for the development of services, wider comparisons of clinical performance made in the absence of any context were considered unreliable and fundamentally misleading.

The ability of the dataset to reflect clinical practice

Many of the physicians' concerns reflect that the dataset was originally designed for administrative use and has never been modified to accommodate the wider purposes for which it is now being used. In particular, specialty codes no longer describe

---

Table 3. Reasons for levels of confidence in routine data to accurately reflect clinical activity.

How confident are you that data routinely coded for central returns accurately reflects your clinical activity?

Reason for level of confidence[*] (with examples)	No opinion	No confidence	Not particularly confident	Reasonably confident	Confident	Very confident
Instinct (Anecdotal evidence; general 'feel' for quality of data)	0	15	15	11	1	0
Prior experience (Audits of data quality; comparison against other data sources)	1	186	186	22	1	0
Engagement with process (Good communication with trust information department or coding staff; clinician coding or validation of coding)	1	1	2	62	19	13
Lack of engagement (Unfamiliarity with process; lack of data provision; poor communication with trust information department)	16	67	91	4	0	2
Dataset limitations (Dataset does not reflect clinical practice due to complexity of clinical problem, responsibility for care or team working; lack of outpatient data)	1	126	107	7	1	3
Coding process (Poor quality discharge summaries; coding staff not clinically trained/under-resourced; general lack of confidence in coding accuracy)	0	119	155	14	4	0
Allocation of activity (Activity allocated to wrong consultant; poorly documented transfers of care)	0	100	91	4	0	0

*Figures comprising participants from England, Wales and Scotland. Many participants provided more than one reason.

Table 4a. Between-group comparisons of responses to post-intervention questions.

Question	Response	Intervention n=47 (%)	Control n=53 (%)	eSCRIPS n=26 (%)
How user friendly did you find the method (you) used to obtain and present this information?	Very unfriendly	2 (4)	4 (8)	0 (0)
	Unfriendly	0 (0)	18 (34)	6 (23)
	Friendly	22 (47)	27 (51)	19 (73)
	Very friendly	23 (49)	3 (6)	1 (4)
	p value		<0.001	<0.001
How has the visit/task affected your view of the usefulness of HES/PEDW/SMR01 data?	Less useful	3 (6)	8 (15)	6 (23)
	No effect	8 (17)	31 (58)	14 (54)
	More useful	35 (75)	5 (9)	5 (19)
	p value		<0.001	<0.001
Will you use the information (you) obtained for your next consultant appraisal?	Yes	32 (68)	41 (77)	21 (81)
	Maybe	10 (21)	10 (19)	1 (4)
	No	5 (11)	2 (4)	4 (15)
	p value		=0.246	=0.377
If yes or maybe, how much of it do you think you will use?	Some	22 (47)	13 (25)	6 (23)
	Most	12 (26)	18 (34)	11 (42)
	All	6 (13)	13 (25)	3 (12)
	p value		<0.05	=0.146
How has the visit/task affected your awareness of the processes involved in data collection?	Not affected	3 (6)	30 (57)	11 (42)
	To some extent	19 (40)	17 (32)	11 (42)
	Significantly	25 (53)	6 (11)	4 (15)
	p value		<0.001	<0.001
To what extent has the visit/task affected your willingness to contribute to the data collection/validation process at a local level?	Less willing	0 (0)	4 (8)	1 (4)
	No effect	5 (11)	28 (53)	14 (54)
	More willing	42 (89)	21 (40)	11 (42)
	p value		<0.001	<0.001
Following the visit/task, do you intend to contact your local information/coding department concerning activity data held in your name?	Yes	29 (62)	10 (19)	4 (15)
	No	2 (4)	20 (38)	12 (46)
	I have already	15 (32)	23 (43)	10 (38)
	p value		<0.005	<0.01
Following the visit/task, do you think you are more or less likely to review or monitor the quality of data collected about your practice?	Less likely	0 (0)	2 (4)	1 (4)
	No change	8 (17)	23 (43)	16 (62)
	More likely	39 (83)	28 (53)	9 (35)
	p value		<0.005	<0.001
Following the visit/task, how frequently would you like to see the coded inpatient and/or day case episodes for patients under your care?	Never	0 (0)	2 (4)	1 (4)
	Rarely	2 (4)	2 (4)	3 (12)
	Occasionally	17 (36)	10 (19)	9 (35)
	Quite regularly	23 (49)	28 (53)	11 (42)
	Often	5 (11)	11 (21)	2 (8)
	p value		=0.123	=0.267
Following the visit/task, how confident are you that data routinely coded for central returns accurately reflects your clinical activity?	No confidence	11 (23)	18 (34)	5 (19)
	Not particularly	9 (19)	19 (36)	12 (46)
	Reasonably	25 (53)	15 (28)	9 (35)
	Confident	0 (0)	0 (0)	0 (0)
	Very confident	1 (2)	1 (2)	0 (0)
	p value		<0.05	=0.245
Are there any ways in which you think your participation in this project may change your future practice?	Yes	35 (74)	21 (40)	12 (46)
	No	12 (26)	31 (58)	14 (54)
	p value		<0.005	<0.05

HES = Hospital Episode Statistics database; n = number; p value = Mann-Whitney comparison against intervention group; PEDW = Patient Episode Database for Wales; SMR01 = Scottish Morbidity Record.

accurately enough the diverse responsibilities of individual consultants working in the NHS, many of whom are active within more than one specialty or sub-specialty. Central returns contain little or no data about specialists who work predominantly in outpatient departments – 19% of those who wished to take part in the study were excluded from randomisation for this reason.

Episodes of activity allocated to a single 'responsible' clinician is a much larger issue, particularly when inpatients are being regularly transferred to colleagues for investigation or treatment and when care is shared between consultants. Considerable variability in the volumes of FCEs between hospitals stems from a lack of standards or central guidance for the interpretation of data dictionary transfer rules, an issue which is only likely to be

Table 4b. Within-group comparisons of responses to pre- and post-intervention questions (Wilcoxon Signed-Rank Test). Percentage totals may fall short of 100 where consultants have not provided an answer.

Question	Response	Intervention n=47 (%)		Control n=53 (%)		eSCRIPS n=26 (%)	
		Pre	Post	Pre	Post	Pre	Post
Following the visit/task, how frequently would	Never	18 (38)	0 (0)	24 (39)	2 (4)	12 (35)	1 (4)
you like to see the coded inpatient and/or day	Rarely	19 (41)	2 (4)	22 (36)	2 (4)	12 (35)	3 (12)
case episodes for patients under your care?	Occasionally	7 (15)	17 (36)	8 (13)	10 (19)	8 (24)	9 (35)
	Quite regularly	2 (4)	23 (49)	5 (8)	28 (53)	1 (3)	11 (42)
	Often	1 (2)	5 (11)	2 (4)	11 (21)	1 (3)	2 (8)
	p value		<0.001		<0.001		<0.001
Following the visit/task, how confident are you	No opinion*	2 (4)	0 (0)	2 (3)	0 (0)	1 (3)	0 (0)
that data routinely coded for central returns	No confidence	17 (36)	11 (23)	21 (36)	18 (34)	12 (36)	5 (19)
accurately reflects your clinical activity?	Not particularly	21 (45)	9 (19)	30 (51)	19 (36)	14 (42)	12 (46)
	Reasonably	7 (15)	25 (53)	6 (10)	15 (28)	7 (21)	9 (35)
	Confident	0 (0)	0 (0)	2 (3)	0 (0)	0 (0)	0 (0)
	Very confident	0 (0)	1 (2)	0 (0)	1 (2)	0 (0)	0 (0)
	p value		<0.05		=0.207		=0.157

*Responses of 'no opinion' were excluded from the Wilcoxon Signed-Rank Test.

Table 5. Categories of data quality issues recorded during consultant visits to *i*Lab.

A   Accuracy

Is the information free from error and inaccuracy?
- Consultant code inaccuracies
- Length of stay inaccuracies
- Waiting list figure inaccuracies

B   Completeness and coverage

Is the information complete? Does is reflect all the activity carried out by staff?
- Lack of outpatient data
- No recording of activity of non-consultant clinicians
- No recording of (inter-specialty) ward referrals and requests for consultation
- No recording of 'ward attender' activity
- No recording of community activity

C   Validity

Are the data items valid? Is the information 'within range' of that expected?
- Invalid or incomplete coding
- Clinically perceived inappropriate coding of primary diagnoses/procedures

D   Timeliness

Is the information available at the right time?
- Delayed 'annual refresh' of HES/PEDW data leading to coding anomalies

E   Relevance for purpose

Is the information contextually appropriate in the eyes of the consultant? Does it reflect clinical practice? Is it relevant for supporting the appraisal and revalidation processes?
- Clinically perceived inappropriate use of specialty code(s)
- Clinically perceived inappropriate use of admission method codes
- Inability to reflect team working
- Problems benchmarking against colleagues
- Under-estimation of activity levels
- Over-estimation of activity levels

F   Distorting factors

Are there local factors or likely artefacts which explain apparent anomalies or differences in the data?
- Working patterns
- Changes in local coding practices
- Changes in referral/booking/admission practices
- Changes in clinical practice
- Factors affecting overall length of stay (longer or shorter than expected)

HES = Hospital Episode Statistics database; PEDW = Patient Episode Database for Wales.

augmented by the wider adoption of Payment by Results.[13] This, coupled with the inability to reflect shared care, ward referrals, teamworking or the activity of nurse practitioners, staff grades and juniors makes for an unrealistic view of clinical practice and performance.

How can the situation be improved?

This study demonstrates the advantages of engaging clinicians in the use of hospital data, highlights issues which can be addressed and raises awareness of the strengths and weaknesses of hospital data. Clinicians will benefit from access to the activity data held in their name, but they need professional and personal support to interpret it. The *i*Lab approach is resource intensive, however, and was not designed to be applied widely. We believe the way ahead is to promote greater clinical engagement through NHS trust information departments, who should be assisted in providing clinically relevant activity data for consultants. This will require recognition by trust management that supporting the information needs of clinicians will support those of the organisation, and will develop a mutual understanding of responsibilities between clinicians, clinical coding staff and information analysts. It will also require investment in clinical coding and information staff time to prepare the analyses and share these with clinicians – investment which is consistently highlighted as being long overdue.[10,14]

However, we believe this is not the final solution. The need for organised collection of routine data to allow the valid and reliable assessment of clinical practice has been emphasised by the Chief Medical Officer for England in his proposals for revalidation.[6] But dedicated national audit databases for every specialty will be a difficult and resource intensive goal. Analysing data captured in electronic records during the course of direct patient care must be the ultimate aim, and improvements in paper-based systems are important steps on the road to accurate electronic records.

Conclusions

The potential for local data to be analysed, examined, and improved is considerable. Sharing these data with consultants to support local appraisal and service development will increase confidence, raise awareness and enable clinicians to engage constructively with the management of health information, while in turn improving the quality of aggregated data and the validity of central returns.

Until appropriate systems are in place that more accurately reflect the complexities of clinical care, however, the monitoring of an individual's performance using centrally returned hospital episode statistics is ill-advised.

References

1 NHS Executive. *Central data collections from the NHS.* Leeds: NHS Executive, 1998. www.dh.gov.uk/en/Publicationsandstatistics/Lettersandcirculars/Healthservicecirculars/DH_4002894

2 Department of Health. *The new NHS: modern, dependable.* London: Stationery Office, 1997. www.archive.official-documents.co.uk/document/doh/newnhs/contents.htm

3 NHS Executive. *Improving 1998/99 HES data quality for clinical indicators and high level performance indicators.* Leeds: NHS Executive, 1999.

4 Department of Health. *Reforming NHS financial flows: introducing payment by results.* London: Stationary Office, 2002. www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_4005300

5 Maynard A, Bloor K. Consultant clinical activity is key to improving productivity. *HSJ* 2006;116:18–9.

6 Department of Health. *Good doctors, safer patients: proposals to strengthen the system to assure and improve the performance of doctors and to protect the safety of patients.* London: Stationery Office, 2006. www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_4137232

7 CHKS. www.chks.co.uk/index.php?compass

8 Dr Foster. Clinician outcomes and benchmarking. www.drfosterintelligence.co.uk/managementInformation/CBT/

9 Williams JG, Mann RY. Hospital episode statistics: time for clinicians to get involved? *Clin Med* 2002;2:34–7.

10 Audit Commission. *Early lessons from payment by results.* London: Audit Commission, 2005. www.audit-commission.gov.uk/reports/national-report.asp?categoryid=&prodid=b502f0fc-e007-4925-ad24-529c4889ad02&fromreportsanddata=national-report&page=index.asp&area=hplink

11 Audit Commission. *Information and data quality in the NHS: key messages from the three years of independent review.* London: Audit Commission, 2004. www.audit-commission.gov.uk/reports/national-report.asp?categoryid=&prodid=4d598af6-3894-401d-aa48-1076125da38d

12 SD2HES section lead and Northgate Information Solutions staff, personal communication, 2005.

13 Audit Commission. *Payment by Results assurance framework: pilot results and recommendations..* London: Audit Commission, 2007. www.audit-commission.gov.uk/reports/national-report.asp?categoryid=&prodid=be765294-44c1-4a5d-be76-1328bcd45747

14 Kennedy I. *The report of the public inquiry into children's heart surgery at the Bristol Royal Infirmary 1984–1995: Learning from Bristol.* www.bristol-inquiry.org.uk/index.htm