AAA benchmarking by Dr Foster: a cause for concern?

VA PANDEY, MI KERLE, MP JENKINS, JHN WOLFE

Regional Vascular Unit, St Mary's Hospital, London, UK

ABSTRACT

INTRODUCTION The publication of interpretable performance data for hospitals is an important service. In November 2002, the medical benchmarking company Dr Foster published a league table based on the results of abdominal aortic aneurysm (AAA) repair. The purpose of our study was to establish the validity of the data used in benchmarking.

PATIENTS AND METHODS Data on elective infra-renal AAA (IRAAA) repair was obtained from three sources. Data used by Dr Foster was based on the hospital PAS system. The databases for both Dr Foster and PAS were analysed and cross-referenced to the vascular unit database maintained by a separately employed audit co-ordinator.

RESULTS Of 395 total aortic aneurysm repairs, 223 (56%) were identified as elective IRAAA repairs on the unit database. Of these, 125 were identified on the PAS database and 115 on the Dr Foster database. The number of deaths was the same in both the unit and Dr Foster databases ($n = 11$) but the Dr Foster database included deaths in patients who had undergone juxtarenal ($n = 1$), Type III TAAA ($n = 2$) and Type IV TAAA ($n = 4$) repairs and omitted 7 deaths following IRAAA. The sensitivity and specificity for the PAS dataset was 0.51 and 0.93, respectively. For Dr Foster, the results were worse with a sensitivity and specificity of 0.41 and 0.86, respectively. The accuracy of the data was 0.6 and 0.69 for Dr Foster and PAS, respectively. Standardised mortality ratios (SMRs) were used to rank hospitals. Dr Foster's published SMR for elective AAA repair for our unit was 160. The actual SMR was 67.

CONCLUSIONS Robust and accurate published league tables should be supported and commended but currently available data appear to be misleading and may cause unnecessary concern to patients.

KEYWORDS

Abdominal aortic aneurysm – League table – Dr Foster

CORRESPONDENCE TO

JHN Wolfe, Consultant Vascular Surgeon, Regional Vascular Unit, 2nd Floor Mary Stanford Wing, St Mary's Hospital, Praed Street, London W2 1NY, UK. E: john.wolfe@st-marys.nhs.uk

Following events leading to the Bristol inquiry, we are in an era of increasing public demand for quality control in hospital medicine.¹ Hospital league tables are a useful tool and allow public scrutiny of hospital performance. Despite these having a high impact in the press and being of great interest to the public, their use remains controversial.^{2–10}

The Dr Foster organisation was established in 2000 as an independent group providing 'authoritative and independent guides to UK health services in the public and private sectors'. In particular, Dr Foster publishes standardised mortality rates for a range of diagnoses and procedures for NHS acute trusts in England, with the aim of allowing consumers and doctors to 'make the best possible choices'.¹¹

In November 2002, Dr Foster published a league table of mortality ratios following repair of abdominal aortic aneurysms (AAA) in *The Times*.¹² These ratios were based on an analysis of the Hospital Episode Statistics (HES),¹³ derived from hospital Patient Administration Systems (PASs), for the period April 1996 to March 2001 inclusive.

We aimed to analyse the elective infra-renal AAA repair dataset used by Dr Foster analysis. We also analysed the source database used by Dr Foster in the analysis, that is the hospital's PAS. We assessed the sensitivity, specificity and accuracy of both datasets by cross-referencing the two datasets with a database prospectively maintained on the Regional Vascular Unit, assuming the data on this database to be accurate.

Patients and Methods

The hospital database for St Mary's Hospital (PAS) provides the basis for statistics forwarded to the UK Department of Health. These data are loaded in a central department according to the Office of Population and Census Statistics (OPCS) operative codes obtained from the patient's discharge sheet.

The data provided to Dr Foster by an individual centre were as follows: gender, date of admission, diagnosis and procedure codes, date of discharge, discharge status and the patient's postcode. The definition of 'episode' was the duration

Table 1 Total elective aneurysm repairs, April 1996 to March 2001

IRAAA	223
JRAAA	33
TAAA (I)	20
TAAA (II)	25
TAAA (III)	35
TAAA (IV)	58
DTAAA	1
Total	395

of care under a particular consultant. Their definition of 'spell' was the entire period in hospital. From these data, Dr Foster sought to include all patient records for which there was: 'a finished episode and spell, valid age, valid gender, non-duplicate records and a valid length of stay'.

Their definition of 'valid' age excluded patients under 45 years. A 'valid length of stay' was no greater than 365 days and any duplicate patient case notes were excluded. Patients who were admitted through inter-hospital transfer (admission code 81) were excluded. Dr Foster included 'spells' (entire admission periods) in which the operation took place in the first 'episode' (*i.e.* 'the patient was admitted for this operation'). The outcome was taken as the outcome of the 'spell', not the episode; patients who died while still in hospital, but after transfer to another consultant, were included.

Dr Foster used OPCS procedure codes L193, L194, L195, L196, L213, L214, L215 and L216 to identify elective AAA repair patients. These codes would 'include bypass, replacement and revision operations' but 'exclude codes which described more complex operations', which may also have included repair of AAA.

The Vascular Unit has an independent, prospectively collected database, the additions to which are validated and discussed on a monthly basis. We, therefore, sought to identify areas of agreement and disagreement between the Vascular Unit database, the hospital database and the Dr Foster analysis. We used the OPCS codes used by Dr Foster (L193, L194, L195, L196, L213, L214, L215 and L216) to identify elective infra-renal abdominal aortic aneurysm (IRAAA) repairs and excluded L193 (for supra-renal aortic repairs). We undertook a separate analysis of a broader group that included patients that had undergone elective juxtarenal AAA (JRAAA) and IRAAA repairs (we included the procedure code L193 on the assumption that Dr Foster also included this code to look at the outcome following JRAAA repair. (Thoraco-abdominal aneurysm [TAAA]

repairs being excluded with Dr Foster's exclusion criteria).

The accompanying report from Dr Foster referred to analysis of 'abdominal aortic aneurysm repair' and did not state whether endovascular repairs were to be included along with open surgical repair. Moreover, three patients in the Dr Foster dataset had undergone endovascular repair of their aneurysm. We, therefore, included the code for endovascular aortic repairs (L26.8 and Y02.1) in addition to the specified codes for open AAA repairs in our search on PAS.

Dr Foster's analysis of mortality rate was standardised for age, gender, length of stay and method of admission using the methodology described by Jarman *et al.*¹⁴ This provided the basis for the ranking of hospitals. We compared Dr Foster's published standardised mortality rate (SMR) for elective AAA repairs with that from the unit database.

Results

The unit's database recorded a total of 395 elective aortic aneurysm procedures performed on patients who were discharged or died between April 1996 and March 2001. Elective IRAAA repair accounted for 223 of these (Table 1). Fourteen of the unit's elective IRAAA patients had undergone endovascular stent repairs. All patients were over 45 years old.

Dr Foster

The Dr Foster dataset included 115 patients, 91 patients also on the Unit Elective IRAAA database; 24 were not correctly identified (Table 2). Fifteen patients had Type IV thoraco-abdominal aneurysm (TAAA) repair.⁵ Patients had juxtarenal aneurysm (JRAAA) repairs. One patient had a type III and one a type II TAAA repair. One of the patients included in Dr Foster's database had not undergone any procedure.

PAS correctly identified 113 patients, with 12 patients incorrectly identified. Five had TAAA repairs, six JRAAA repairs and one included on Dr Foster's database, was the patient that had no operation. One patient had been given the inter-hospital transfer code (81) on PAS. All PAS patients

Table 2 Dr Foster's elective AAA dataset

IRAAA	91
TAAA (IV)	15
JRAAA	5
TAAA (III)	2
TAAA (II)	1
No operation	1
Total	115

had completed 'spells', non-duplicate records (patients previously seen in the hospital may have another hospital number and set of medical case notes) and length of stay. No difference was found between the three databases in identifying gender or date of birth. Three of the IRAAA repairs evaluated by Dr Foster were endovascular aneurysm repairs.

Sensitivity, specificity and accuracy

Using the unit database as the gold standard, the sensitivity and specificity for the Dr Foster dataset was 0.41 and 0.86 and for PAS 0.51 and 0.93, respectively (assuming the prospectively collected unit database to be accurate). The accuracy for the Dr Foster and PAS datasets were 0.6 and 0.69, respectively

Distribution of deaths and calculation of standardised mortality rate

The unit database recorded 11 deaths in the 223 elective AAA patients. Dr Foster also recorded 11 deaths in its group of 115 AAA patients; however, seven deaths listed in Dr Foster's database involved patients who had not undergone IRAAA repair. The 11 deaths are highlighted in Table 3.

Dr Foster's published Standardised Mortality Rates (SMRs) for elective AAA repair for our unit was 160 (100 being the mean). Using the mortality rate from the unit database, the SMR was 67 (assuming the data from other hospitals are unchanged).

Juxtarenal abdominal aortic aneurysm repair

If analysis is based on an elective AAA group that includes both JRAAA and IRAAA repairs, the level of discrepancy in SMR increases. The unit database identified 33 patients who had undergone elective JRAAA repair between April 1996 and March 2001. All were over 45 years old. No patient had been given admission code 81 for inter-hospital transfers on PAS. Two of the 33 patients who had undergone elective JRAAA repair died within the identified period. Combining IRAAA and JRAAA patients in the group for analysis would then give a total of 256 patients on the unit database, with 13 deaths. The Dr Foster database identified 96, with 19 incorrect inclusions. PAS included 116, with 26 incorrect inclusions. Our SMR for juxta- and infrarenal aneurysms would therefore be 69, as opposed to the published 160.

Analysis

Examination of the 110 patients on the unit elective IRAAA database not identified by PAS showed three comparably sized groups:

CODED AS EMERGENCY ADMISSION

Thirty-seven patients had been given emergency admission codes (21, 22, 23, 24 or 28). These patients may have been admitted as emergencies but had their operation on an

Table 3 Distribution of deaths from IRAAA repair on Dr Foster database

TAAA (IV)	4
IRAAA	4
TAAA (III)	2
JRAAA	1
Total	11

elective basis after a period of pre-operative investigation and optimisation. The National Vascular Database¹⁵ takes into account mode of admission as well as timing of surgery. The Dr Foster methodology failed to take this into account.

NO PRIMARY PROCEDURE IN FIRST EPISODE

The criteria for inclusion into the Dr Foster database were 'AAA repair in the first hospital episode'. Some patients (and this is not consistent) have a vascular admission subdivided into two episodes – the first episode may be an investigation and assessment episode, the operation taking part in the second episode. Forty-two patients had no primary procedure recorded in their first episode. By applying the Dr Foster inclusion criteria, these patients would have been omitted from its database.

ERRORS IN CODING

The remaining 31 patients had either an investigative or radiological procedure as their primary procedure, or had their operative procedure miscoded. This was confirmed by a review of case notes.

Discussion

The publication of interpretable results for public scrutiny is a potentially important service. Hitherto, some of the information presented in the press has been criticised by the profession, both for its sensationalism and its inaccuracy. The data collection has been insufficiently sophisticated to assess areas of concern accurately. Dr Foster, therefore, set out to provide an independent and accurate assessment of performance. When *The Times* published the results of infra-renal abdominal aortic aneurysm repair, the results for St Mary's Hospital were unexpected. We, therefore, sought to identify and assess the reason for any discrepancies.

Forty-two patients on the unit elective IRAAA database had no primary procedure recorded in their first admission episode on PAS. However, 26 of these had an aortic procedure recorded as a second episode. Dr Foster's analysis does not take into account the continuing use of 'completed

consultant episodes' as the National Health Service measure of hospital activity. Furthermore, 31 patients having IRAAA and 12 patients having JRAAA were given a diagnostic or radiological procedure as their primary procedure, or had their operative procedure miscoded.

There are also vague codes (*e.g.* L195 'Replacement of aneurysm segment of abdominal aorta by anastomosis of aorta to aorta'). The availability of these codes makes classification problems inevitable. In particular, it undermines the effectiveness of 'adjustment made in logistic regression model for position of repair: suprarenal and infrarenal', as carried out by Dr Foster in the subsequent year's data analysis.

Dr Foster also aimed to 'exclude inter-hospital transfers' from analysis by basing exclusion on admission code 81 'Transfer of any admitted patient from other hospital provider other than in an emergency'. One patient had been given this admission code. Twenty-six patients had been transferred from other hospitals and given admission code 28 'Admission by...other means, including admission from the accident and emergency department of another provider where they had not been admitted'.

We were uncertain whether endovascular repair of AAA should be included; however, as the Dr Foster database listed three patients who had endovascular procedures, we included the code for endovascular repair of aneurysms (L26.8 and Y02.1). Since endovascular repairs accounted for only 14 cases in the unit database, it does not explain the differences between the Dr Foster and unit databases.

Analysis of the status of patients on the unit's elective AAA database but 'missing' from the Dr Foster dataset showed three categories of omission: (i) to admission code; (ii) if the patient's operation did not occur in the first 'episode'; and (iii) if the operation was miscoded.

Michaels¹⁶ predicted the above as potential sources of miscoding with Dr Foster's data. He suggests methods by which these errors may be reduced including aggregation of all aortic procedures with a definition of emergency and elective groups based on the method of admission to hospital as well as aggregation of all aortic codes to consider death after all aortic surgery (*i.e.* aorto-bifemoral bypass for occlusive disease) rather than aneurysm surgery alone.

League tables of crude death rates may be misleading, and McKee and Hunter⁷ suggested that any decision to publish such tables is based on political, rather than scientific, grounds. Some adjustment for illness severity should be made.^{6,9,17-19} These adjustments may not allow for differences in hospital volume that has been shown to be inversely proportional to mortality.²⁰⁻²⁴

Jacobson *et al.*⁴ questioned league tables and raised concerns that their publication may lead to unintended changes in behaviour. This has been seen in the US with fewer cardiac surgeons willing to take on high-risk cases.²⁵ They also discuss the findings in New York following 'report cards' on cardiac surgeons. A

3-fold increase in reporting of chronic obstructive pulmonary disease and a 4-fold increase in congestive cardiac failure effectively served to reduce severity-adjusted mortality rates.²⁶

League tables may not be the most effective way to deliver performance data for hospitals. Mohammed *et al.*⁸ suggested that the use of mortality control charts would have identified a poorly performing cardiac surgery unit in Bristol 11 years before external action to address concerns was undertaken. Tekkis *et al.*²⁷ also support the use of mortality control charts for gastro-oesophageal cancer surgery to identify hospitals where performance diverges significantly from the mean.

Well-constructed and robust performance tables are to be commended and their open availability has been recommended in the National Performance Framework. We have shown that the collection of data needs to be improved in order to avoid misinformation.

First, the profession has a responsibility to ensure accuracy of all recorded data. This can be achieved by surgeons coding operations personally either whilst in theatre (as is the case in some hospitals) or by closer co-operation with hospital coding staff. Vague OPCS codes relating to undefined procedures are unavoidable in vascular surgery and such data would need to be entered elsewhere. Errors relating to the hospital 'episode' the operation took place on (for example, if a patient underwent pre-operative investigations in the first 'episode') would also be reduced by communication between coding staff and the operating team.

Second, it is the responsibility of hospitals to ensure collection and distribution of accurate data to the Department of Health. This can be achieved with dedicated audit staff working for the unit or hospital to ensure accurate collection of data for the hospital's PAS. Well-designed, prospectively maintained databases may be particularly useful in this respect, as it would be easy to collect uniform data on a national level. The Vascular Society of Great Britain and Ireland run a National Vascular Database²⁸ that collects data for index operations including abdominal aortic aneurysm repair. Participation is voluntary but the data are collected, validated and submitted by surgical units rather than filtered from potentially inaccurate PASs and surgeons are responsible for all submitted data. This is an alternative source for Dr Foster and the Department of Health. The success of a national project such as this depends on universal participation by vascular surgeons.

Lastly, it is the responsibility of organisations such as Dr Foster to ensure validity of the data used prior to publicising results that may cause unnecessary concern to patients and referring practitioners.

Acknowledgement

This work was presented, in part, at the Annual Scientific Meeting of the Vascular Surgical Society of Great Britain and Ireland, November 2003.

References

1. Smith R. All changed, changed utterly. British medicine will be transformed by the Bristol case. *BMJ* 1998; **316**: 1917–8.
2. Aylin P, Jarman SB, Kelsey T. What hospital mortality league tables tell you. *BMJ* 2003; **326**: 1397–8.
3. Dean M. Hospital mortality league tables. *Lancet* 1994; **344**: 1561.
4. Jacobson B, Mindell J, McKee M. Hospital mortality league tables. *BMJ* 2003; **326**: 777–8.
5. Joyce R, Webb R, Peacock JL, Stirland H. Adjusted mortality rates: a tool for creating more meaningful league tables for stillbirth and infant mortality rates. *Public Health* 2002; **116**: 315–21.
6. Marshall EC, Spiegelhalter DJ. Reliability of league tables of *in vitro* fertilisation clinics: retrospective analysis of live birth rates. *BMJ* 1998; **316**: 1701–4, discussion 1705.
7. McKee M, Hunter D. Mortality league tables: do they inform or mislead? *Qual Health Care* 1995; **4**: 5–12.
8. Mohammed MA, Cheng KK, Rouse A, Marshall T. Bristol, Shipman, and clinical governance: Shewhart's forgotten lessons. *Lancet* 2001; **357**: 463–7.
9. Parry GJ, Gould CR, McCabe CJ, Tarnow-Mordi WO. Annual league tables of mortality in neonatal intensive care units: longitudinal study. International Neonatal Network and the Scottish Neonatal Consultants and Nurses Collaborative Study Group. *BMJ* 1998; **316**: 1931–5.
10. Winston R. League tables of *in vitro* fertilisation clinics misinform patients. *BMJ* 1998; **317**: 1593–4.
11. Dr Foster – Your guide to better health. <www.drfooster.co.uk> accessed January 2006.
12. Hawkes N. Artery operations four times riskier in some hospitals. *The Times* 18 November 2002; 6–7.
13. Department of Health. *Hospital Episode Statistics*. London: Department of Health <www.dh.gov.uk> accessed January 2006.
14. Jarman B, Gault S, Alves B, Hider A, Dolan S, Cook A *et al*. Explaining differences in English hospital death rates using routinely collected data. *BMJ* 1999; **318**: 1515–20.
15. Earnshaw JJ. Clinical outcomes audit in vascular surgery: a shield for our profession. *Ann R Coll Surg Engl* 2003; **85**: 256–9.
16. Michaels JA. Use of mortality rate after aortic surgery as a performance indicator. *Br J Surg* 2003; **90**: 827–31.
17. Dubois RW, Rogers WH, Moxley 3rd JH, Draper D, Brook RH. Hospital inpatient mortality. Is it a predictor of quality? *N Engl J Med* 1987; **317**: 1674–80.
18. Selker HP. Systems for comparing actual and predicted mortality rates: characteristics to promote cooperation in improving hospital care. *Ann Intern Med* 1993; **118**: 820–2.
19. Hartz AJ, Krakauer H, Kuhn EM, Young M, Jacobsen SJ, Gay G *et al*. Hospital characteristics and mortality rates. *N Engl J Med* 1989; **321**: 1720–5.
20. Birkmeyer JD, Warshaw AL, Finlayson SR, Grove MR, Tosteson AN. Relationship between hospital volume and late survival after pancreaticoduodenectomy. *Surgery* 1999; **126**: 178–83.
21. Birkmeyer JD, Finlayson SR, Tosteson AN, Sharp SM, Warshaw AL, Fisher ES. Effect of hospital volume on in-hospital mortality with pancreaticoduodenectomy. *Surgery* 1999; **125**: 250–6.
22. Birkmeyer JD. Relation of surgical volume to outcome. *Ann Surg* 2000; **232**: 724–5.
23. Birkmeyer JD, Finlayson EV, Birkmeyer CM. Volume standards for high-risk surgical procedures: potential benefits of the Leapfrog initiative. *Surgery* 2001; **130**: 415–22.
24. Birkmeyer JD, Siewers AE, Finlayson EV, Stukel TA, Lucas FL, Batista I *et al*. Hospital volume and surgical mortality in the United States. *N Engl J Med* 2002; **346**: 1128–37.
25. Schneider EC, Epstein AM. Influence of cardiac-surgery performance reports on referral practices and access to care. A survey of cardiovascular specialists. *N Engl J Med* 1996; **335**: 251–6.
26. Green J, Wintfeld N. Report cards on cardiac surgeons. Assessing New York State's approach. *N Engl J Med* 1995; **332**: 1229–32.
27. Tekkis PP, McCulloch P, Steger AC, Benjamin IS, Poloniecki JD. Mortality control charts for comparing performance of surgical units: validation study using hospital mortality data. *BMJ* 2003; **326**: 786–8.
28. The Vascular Society of Great Britain and Ireland. *National Vascular Database*. <www.vascularsociety.org.uk> accessed January 2006.