Medical Informatics and the Internet in Medicine
June 2005; 30(2): 151 – 156

Taylor & Francis
Taylor & Francis Group

Measuring clinical performance using routinely collected clinical data

D. R. PRYTHERCH[1], J. S. BRIGGS[1], P. C. WEAVER[1], P. SCHMIDT[1] & G. B. SMITH[2]

[1]*University of Portsmouth, Portsmouth, UK and* [2]*Queen Alexandra Hospital, Portsmouth, UK*

(Received for publication April 2005)

Abstract
Following the well-publicized problems with paediatric cardiac surgery at the Bristol Royal Infirmary, there is wide public interest in measures of hospital performance. The Kennedy report on the BRI events suggested that such measures should be meaningful to the public, case-mix-adjusted, and based on data collected as part of routine clinical care. We have found that it is possible to predict in-hospital mortality (a measure readily understood by the public) using simple routine data—age, mode of admission, sex, and routine blood test results. The clinical data items can be obtained at a single venesection, are commonly collected in the routine care of patients, are already stored on hospital core IT systems, and so place no extra burden on the clinical staff providing care. Such risk models could provide a metric for use in evidence-based clinical performance management. National application is logistically feasible.

Keywords: *Clinical performance, clinical data, clinical outcome modelling, hospital performance*

1. Introduction

The events in Bristol between 1984 and 1995 focused Government, media, and public attention on the performance of clinicians, clinical departments, and hospitals in the UK. The Department of Health (DH) [1] and others (e.g. Dr Foster [2], Sunday Times Good Hospital Guide [3]) now publish comparative performance data for hospitals based on Hospital Episode Statistics (HES).

The Kennedy Report on the Bristol enquiry stated [4]:

> Hospitals and the NHS could tell you about throughput, bed occupancy and the costs involved. But, generally speaking, the quality of outcome was a closed book.

Correspondence: D. R. Prytherch, Healthcare Computing Group, Department of Information Systems and Computer Applications, University of Portsmouth, 1 Burnaby Road, Portsmouth, Hants, PO1 3AE, UK. Tel: +44-23-9284 6231. E-mail: dave.prytherch@port.ac.uk

ISSN 1463-9238 print/ISSN 1464-5238 online © 2005 Taylor & Francis
DOI: 10.1080/14639230500298966

> At national level, the indicators of performance should be comprehensible to the public as well as to healthcare professionals. They should be fewer and of high quality, rather than numerous but of questionable or variable quality.
>
> Variables such as case mix ... must be allowed for, so that, wherever feasible, it is possible to compare like with like.
>
> For the future the multiple methods and systems for collecting data must be reduced. Data must be collected as the by-product of clinical care.

Kennedy recommended that HES be '... supported as a major national resource which can be used reliably, with care, to undertake the monitoring of a range of healthcare outcomes'. In its response, the DH stated, 'They do not contain the necessary clinical details to make full allowance for case mix ... and so cannot be used to produce absolute measures of quality' [5]. HES, and analyses based on them alone, have little clinical acceptance since they are not case-mix-adjusted.

Hospitals, such as Portsmouth Hospitals, have large databases used in the operational care of individual patients, but typically, these are not exploited in an aggregated manner.

2. Case-mix adjustment

The terms *case-mix adjustment*, *risk adjustment* and *risk stratification* are effectively synonymous. Appropriate case-mix adjustment is essential to compare performance between hospitals, clinical groupings or individual clinicians, or to investigate their performance over time. Case-mix adjustment, by definition, adjusts for differences in patient populations to allow such comparisons. Case-mix adjustment is necessary (but not of itself sufficient) to allow appropriate comparison between different providers of care. Organizational, resource and process of care differences are of great importance [6]. However, case-mix adjustment will help to expose the effects of such differences—which are more amenable to improvement by clinicians and managers than case mix.

Case-mix adjustment is necessarily limited and is never complete, as one can only risk adjust within the dimensions of the data items in the risk model. For complete case-mix adjustment, one would need to know everything about all patients—a large if not infinite number of data items, many of which are currently unknown. In practice, one must be guided by engineering realities and attempt to describe the population, if not every individual within it, with as small a data set as possible. Care must be taken with the data items included in the risk model. The inclusion of process- or resource-orientated data items may obfuscate rather than highlight the effect of these on clinical outcomes. The purpose of a particular risk model must be understood.

Case-mix adjustment usually involves the creation of mathematical models, using relevant patient data, that predict a patient's risk of suffering an adverse clinical outcome, typically mortality. These predictions are then aggregated for the patient group(s) being studied and compared with those observed. Such models predict risk of adverse outcome for each patient and thus stratify risk; they do not predict which patients will suffer adverse events. The ability to predict expected clinical outcomes provides a ruler to measure performance and the effects of change (process, resource, etc.), and can highlight exceptionally good or poor performance for further investigation or to promote the dissemination of good practice [6,7].

3. Existing work

The clinical performance measures produced by Dr Foster (and also published on their Web site and in the Times and Sunday Times), CHKS, and the Department of Health use statistical models of clinical outcomes based on coded administrative data. This approach, and the uses to which it is put, has generated considerable reservations in the medical literature [6,8,9]. These are largely based on concerns regarding the absence of case-mix adjustment in the models, the accuracy of coded administrative data, and the potential for manipulation (or gaming) of these administrative data items to improve the apparent performance of provider organizations.

Surgeons in the UK have responded actively to the challenges raised by scrutiny—e.g. the Vascular Surgical Society of Great Britain and Ireland and the Society of Cardiothoracic Surgeons of Great Britain and Ireland publish annual reports [10,11] based in large part on case-mix-adjusted risk models of adverse clinical outcomes.

In the US, considerable work on modelling outcomes following surgery has been performed for the Department of Veterans Affairs by the Khuri group [12] and their risk models have been used as part of a programme to raise clinical standards [7]. In the UK, POSSUM (Physiological and Operative Severity Scoring in the enUmeration of Morbidity and Mortality) and its Portsmouth-POSSUM based variants [13–16] have been developed. These models require the collection of considerable data, up to 25 items in the case of P-POSSUM and 34 to 55 items in the Khuri work. The P-POSSUM methodology has been used internationally [17–19] and was used in a recently published comparison of UK and US surgical outcomes [20].

There is an inevitable tension between the collection of these data and the provision of care. However well crafted, data-collection processes are most likely to fail for high-risk emergency admissions, particularly at night, when collection of a large number of data items may not be logistically feasible within the clinical context. If problems of selection are to be avoided, the necessary data must be collected for all patients.

Additionally, the lack of functionality of current hospital information technology (IT) systems has meant there was anyway no computer repository available to collect these data on a routine basis.

4. BHOM: Biochemistry and haematology outcome models

To overcome these problems, we investigated modelling adverse clinical outcomes using only data stored in core hospital systems at Portsmouth Hospitals, one of the largest acute hospitals in England. These were limited to the patient administration system (PAS) and the biochemistry and haematology elements of the pathology database. To date we have avoided the use of coded administrative data because of the, perhaps overstated, reservations regarding these in the literature [8,9]. Our choice of clinical data items was determined by those most commonly available for the specialities investigated, and were full blood count, urea, and electrolytes, and serum albumen. The administrative data items used are age at admission, sex, and mode of admission—that is, elective or emergency.

To date, the specialities successfully modelled account for over 80% of the in-patient mortality. It has been possible to form good models in all the specialities investigated so far. The models are formed using binary logistic regression following the methods of Hosmer and Lemeshow [21]. All models are tested by prospective application against further data sets—a more rigorous test than assessing the model against the data used to form the model.

To illustrate the predictive performance of the technique, the results obtained for mortality at discharge in general surgery are summarized below. This has been published in full [22]. Separate models were required for elective admissions undergoing operation, emergency admissions undergoing operation, and emergency admissions that did not undergo operation (there were too few non-operative elective admissions to allow modelling). For the operative admissions, the models included a measure of the complexity and extent of the operation based on the BUPA (British United Provident Association) operative severity score, widely used in the private medical sector in the UK. This has been shown to be a useful measure of surgical workload and complexity of operation [23]. The study covered the 5 year period, 1st August 1997 to 31st July 2002. There were some 28,925 general surgical in-patient episodes with necessary data. Models were constructed from years 1 and 2, and tested prospectively against years 3, 4, and 5. Table I summarizes the combined performance of the three models in the final 6 month period of the study—some 3 years after the data that gave rise to the models.

For each episode, the appropriate model produces a predicted risk of death at discharge. Using the predicted risk of death, the episodes are grouped into risk ranges chosen to be clinically useful, give similar predicted deaths in each range, and ensure at least five predicted deaths in each range. For each risk range, the mean risk is calculated, and hence the number of predicted deaths in each risk range can be calculated. This is compared with the number of reported deaths. Calibration, or goodness of fit, is assessed using the χ^2 test. This is a null hypothesis test—models with p values greater than 0.05 are considered to show no evidence of lack of fit, and vice versa.

5. Summary

BHOM gives good accuracy as measured by calibration and discrimination. The clinical data can be obtained from a single venesection, are logistically feasible to collect for all patients, are collected as part of routine care, and so place no extra burden on those providing care. All data are objective and subject to quality control by the pathology laboratory. The approach is resistant to inappropriate manipulation (gaming). The data are clinically meaningful and trusted by clinical staff. The data for predicting risk are available very soon after admission.

The results presented here were for general surgery. We have recently published the results of applying the BHOM approach to general medicine patients [24]. The BHOM approach successfully modelled mortality at discharge in general medicine.

Existing monitoring and surveillance systems [1,2] require coded administrative data only available after discharge. Our techniques add clinical context to these and have obvious uses

Table I. Demonstration of the prospective performance of the combined models for general surgery[a].

Risk range (%)	Episodes	Mean risk (%)	Predicted deaths	Reported deaths	χ^2
$\geqslant 0$ to $\leqslant 5$	2227	0.98	22	24	0.23
>5 to $\leqslant 10$	261	7.09	19	21	0.36
>10 to $\leqslant 15$	119	12.13	14	14	0.01
>15 to $\leqslant 25$	91	19.13	17	20	0.48
>25 to $\leqslant 50$	59	34.33	20	19	0.12
>50 to 100	24	64.17	15	14	0.36
$0\geqslant$ to 100	2751	3.91	108	112	1.56

[a]$\chi^2 = 1.56$, 6 d.f., $p = 0.96$, no evidence of lack of fit.

in clinical governance and clinical performance management. Our approach only uses data available immediately after admission.

BHOM has been successfully applied to the National Vascular Database [25], giving us considerable confidence that our approach is generally applicable. However, it is essential that, before it is used to influence patient care, it should be validated in more than one hospital for as wide a range of specialities and operative and diagnostic groups as possible.

This work shows that if simple pathology data are added to HES, case-mix adjustment can be achieved. In the UK NHS context, this would allow HES to become the national resource to provide measures of clinical performance, acceptable to clinicians, envisaged by Kennedy. All hospitals have PAS and pathology systems, but at present HES data are only collected from PAS. It seems a simple extension to add data to HES from a second hospital core system.

Performance management is with us but is currently bedevilled by the lack of case-mix adjustment. The NHS is expending considerable resources on IT systems though the National Programme for IT. It is to be hoped that these will contain appropriate clinical data to enable case-mix adjustment and so facilitate evidence-based performance management using clinically based and relevant information. We consider this particularly important at a time when the NHS is actively investigating different models of care—it is surely desirable that judgements regarding models of care should take appropriate account of the clinical outcomes of the care.

Acknowledgements

Sources of financial support were Portsmouth NHS Research and Development Consortium, Portsmouth Hospitals NHS Trust and the University of Portsmouth.

References

1. Department of Health. NHS performance indicators: National figures: February 2002. 1st Feb 2002. Available from: http://www.dh.gov.uk/assetRoot/04/05/96/41/04059641.pdf (accessed 13th October 2005).
2. Available from: www.drfoster.co.uk.
3. Available from: www.timesonline.co.uk.
4. Learning from Bristol: The report of the public inquiry into children's heart surgery at the Bristol Royal Infirmary 1984–1995. Ian Kennedy, Produced for the Bristol Royal Infirmary Inquiry by COI CommunicationsCM Published by The Stationery Office Limited. Available from: http://www.bristol-inquiry.org.uk/
5. Learning from Bristol: The Department of Health's Response to the Report of the Public Inquiry into children's heart surgery at the Bristol Royal Infirmary 1984–1995. London: The Stationery Office; 2002.
6. Lilford R, Mohammed MA, Spiegelhalter D, Thomson R. Use and misuse of process and outcome data in managing performance of acute medical care: avoiding institutional stigma. Lancet 2004;363:1147–1154.
7. Khuri SF, Daley J, Henderson WG. The comparative assessment and improvement of quality of surgical care in the Department of Veterans Affairs. Arch Surg 2002;137:20–27.
8. McKee M, Coles J, James P. 'Failure to rescue' as a measure of quality of hospital care: the limitations of secondary diagnosis coding in English hospital data. J Public Health Med 1999;21:453–458.
9. Jackson B, Mindell J, McKee M. Hospital mortality league tables: question what they tell you—and how useful they are. British Medical Journal 2003;326:777–778.
10. Ashley S, Ridler B, Kinsman R. Vascular Surgical Society of Great Britain and Ireland National Vascular Database Report 2002. Henley-on-Thames, UK: Dendrite Clinical Systems; 2003.
11. Keogh BE, Kinsman R. The Society of Cardiothoracic Surgeons of Great Britain and Ireland National Adult Cardiac Surgical Database Report 2000–2001. Henley-on-Thames, UK: Dendrite Clinical Systems; 2002.
12. Khuri SF, Daley J, Henderson W, Hur K, Gibbs JO, Barbour G, Demakis J, Irvin G 3rd, Stremple JF, Grover F, McDonald G, Passaro E Jr, Fabri PJ, Spencer J, Hammermeister K, Aust JB. Risk adjustment of the postoperative mortality rate for the comparative assessment of the quality of surgical care: results of the National Veteran Affairs Surgical Risk Study. Journal of American College of Surgeons 1997;185(4):325–338.

13. Whiteley MS, Prytherch DR, Higgins B, Weaver PC, Prout WG. An evaluation of the POSSUM Surgical Scoring System. British Journal of Surgery 1996;83:812–815.
14. Prytherch DR, Whiteley MS, Higgins B, Weaver PC, Prout WG, Powell SJ. POSSUM and P-POSSUM for predicting mortality. British Journal of Surgery 1998;85:1217–1220.
15. Prytherch D, Sutton GL, Boyle JR. Portsmouth-POSSUM Models for abdominal aortic aneurysm Surgery. British Journal of Surgery 2001;88:958–963.
16. Tekkis PP, McCulloch P, Poloniecki JD, Prytherch DR, Kessaris N, Steger AC. Risk-adjusted prediction of operative mortality in oesophagogastric surgery with O-POSSUM. British Journal of Surgery 2004;91:288–295.
17. Yii K, Ng KJ. Risk-adjusted surgical audit with the POSSUM scoring system in a developing country. British Journal of Surgery 2002;89:110–113.
18. Lam C-M, Fan ST, Yuen AW, Law WL, Poon K. Validation of POSSUM scoring systems for audit of major hepatectomy. British Journal of Surgery 2004;91:450–454.
19. Mohil RS, Bhatnagar D, Bahadur L, Rajneesh, Dev FK, Magan M. POSSUM and P-POSSUM for risk-adjusted audit of patients undergoing emergency laparotomy. British Journal of Surgery 2004;91:500–503.
20. Bennett-Guerrero E, Hyam JA, Shaefi S, Prytherch DR, Sutton GL, Weaver PC, Mythen MG, Grocott MP, Parides MK. Comparison of P-POSSUM risk-adjusted mortality rates after surgery between patients in the United States of America and the United Kingdom. British Journal of Surgery 2003;90:1593–1598.
21. Hosmer DW, Lemeshow S. Applied logistic regression. New York: Wiley; 1989.
22. Prytherch DR, Sirl JS, Weaver PC, Schmidt P, Higgins B, Sutton GL. Towards a national clinical minimum dataset for general surgery. British Journal of Surgery 2003;90:1300–1305.
23. Jones SM, Collins CD. Caseload or workload? Scoring complexity of operative procedures as a means of analysing workload. British Medical Journal 1990;301(6747):324–325.
24. Prytherch DR, Sirl JS, Schmidt P, Featherstone PI, Weaver PC, Smith GB. The use of routine laboratory data to predict in-hospital death in medical admissions. Resuscitation 2005;66:203–207.
25. Prytherch DR, Ridler BMF, Ashley S, on behalf of the Audit and Research Committee of The Vascular Society of Great Britain and Ireland. Risk adjusted predictive models of mortality after index vascular operations using a minimal data set. British Journal of Surgery 2005;92:714–718. Published on line DOI: 100.1002/bjs.4965.