Use of process measures to monitor the quality of clinical practice

Outcomes of care are a blunt instrument for judging performance and should be replaced, say **Richard J Lilford, Celia A Brown, and Jon Nicholl**

Healthcare organisations are increasingly scrutinised by external agencies, such as the Health Care Commission in England and Medicare in the United States. Such agencies increasingly concern themselves with the quality of care and not just measures of throughput, such as waiting times and the average length of hospital stay. Measures of clinical quality are also likely to be used increasingly to monitor the performance of individual doctors.¹ But how should quality be measured? The intuitive response is to measure the outcomes of care—after all, patients use the service to improve their health outcomes. We argue that this beguiling solution has serious disadvantages because of the poor correlation between outcome and quality and that use of outcome as a proxy for quality is a greater problem when the data are used for some purposes than for others.

Purpose of measurement

Data on quality can be used either for internal quality improvement or for external reporting. In the first scenario, data are collected by an organisation or individual for internal audit in the spirit of continuous improvement (quality circles, total quality management, plan do act, Kaizen, etc). In the second scenario, monitoring is imposed externally by health service funders for purposes of accountability (performance management). When results lie above or below some predefined threshold, funders may use the data to prompt further investigation in a completely non-pejorative manner. Alternatively, they may use data as the basis for sanction or reward. For example, hospitals may be given ratings that determine managerial freedoms and financial reward or a doctor may be suspended. We shall refer to use of data for sanction or reward as data for judgment. It is such use that is particularly problematic.

Outcomes and quality

The main disadvantage of measuring outcomes arises from the low signal to noise ratio: outcomes are likely to be affected by factors other than the quality of care. A recent systematic review showed that, although statistically significant, the correlation between the quality of clinical practice and hospital mortality is low² and hence mortality is neither a sensitive nor a specific test for quality. Modelling shows that big differences in the quality of care are likely to be lost in mortality statistics³ and that over half of the institutions with the worst quality of care are likely to have mortality in the normal range and vice versa.⁴ The situation may be worse at the community level.⁵

Richard J Lilford professor of clinical epidemiology, **Celia A Brown** research fellow, Department of Public Health and Epidemiology, University of Birmingham, Birmingham B15 2TT

Jon Nicholl director MCRU policy research programme, School of Health and Related Research, University of Sheffield, Sheffield S14DA

Correspondence to: R J Lilford
r.j.lilford@bham.ac.uk

Accepted: 25 June 2007

It is a myth that the problem of poor correlation between quality and outcomes can be solved by statistical adjustment for risk (the risk adjustment fallacy).⁶ Risk adjustment does not remove the problems of bias in rankings for two reasons:

Firstly, risk adjustment cannot allow for case mix variables that have not been measured (perhaps because they are unknown) and are therefore omitted from the statistical model. Nor can it allow for differences in definitions (or in how the same definitions are applied) to either numerators or denominators. For instance, differences in discharge policies (perhaps influenced by availability of a local hospice) will affect the types of patients included in the statistics.

Secondly, risk adjustment is sensitive to modelling assumptions. Adjustment may even increase bias if the risk associated with the risk factor is not constant across groups being compared.^{7 8} For example, the effect of age on mortality may be greater in some groups (such as those from low socioeconomic backgrounds) than in others. If this is the case risk adjustment will under-adjust for groups in which age has the largest effect. The predicted mortality will be lower than the observed mortality and the playing field is tilted against clinicians or institutions in places where the age effect is greatest.

The problems of risk adjustment do not just apply to mortality but also to other outcomes, such as surgical site infections, for which the definition varies widely from place to place. Using surrogate outcomes, such as the proportion of diabetic patients whose measure of glucose control exceeds some threshold, brings in further confounding variables, such as systematic differences in patients' willingness to adhere to treatment. Even using patient satisfaction to rank individuals or institutions for humane care is potentially misleading since the results may be confounded by systematic differences in expectations.⁹

There may be some topics where the signal to noise ratio is much better than in the examples cited above and hence where a sizeable portion of the variance in outcome is made up of variance in quality. If such examples exist, they are likely to arise among the most technically demanding services such as paediatric cardiac surgery. Claims and counter claims have been made about the role of outcome monitoring for coronary artery bypass surgery.¹⁰ We suspect that even in these topics case selection and other differences between institutions have the major role. We therefore believe those who wish to use such outcomes for performance management within an

accountability framework must first prove that they are strongly correlated (not just statistically associated) with quality.

Using outcomes for sanction or reward

With some possible exceptions, outcomes are clearly neither sensitive nor specific measures of quality. Managers and clinicians therefore quite properly distrust them. This can induce perverse incentives—staff apply their ingenuity to altering or disproving the figures rather than tackling quality or safety, patients are nudged into more severe prognostic categories, treatment may be targeted at patients with the best prognosis (who are often those with the least capacity to benefit), and there are even cases where statutory data have been altered.^{11 12}

The problem that outcome data are poor barometers of clinical quality is viciously confounded by both their inability to discriminate between good and poor performers¹³ and the lack of information they convey about how improvements should be made. In education, for example, it is now standard practice to include a comment and not just a grade when assessing an essay or assignment. Using outcomes to trigger sanctions or rewards may induce a sense of shame or institutional stigma⁶—the feeling of diminished status that comes of being branded bad without being told what the problem is.

When outcomes are used to judge the performance of individual clinicians further problems arise. Firstly, the results are less precise than they are at institutional level. Secondly, outcomes synthesise all of the processes received by the patient and therefore reflect the activities of many clinicians and support services.

Process: an alternative measure

Measures of clinical process have many advantages over outcomes. These advantages are particularly important if policy makers insist on using data for judgment. Clearly, the processes selected for scrutiny must comprise accepted and scientifically valid tenets of clinical care: do patients with a fractured neck of the femur get surgery within 24 hours? Are patients on ventilation nursed in a semi-prone position? Do clinicians monitor respiratory rate on the acute medical wards and, if so, do they respond promptly to signs of deterioration?

Such measures are not a panacea. The measures themselves must be valid and important. Furthermore, process measures are not immune from case mix bias; sicker patients challenge the system more than those who are not so sick, so the playing field is tilted against those who care for more vulnerable patients. Nevertheless, we believe that process measures have four fundamental advantages over outcomes:

Reduction of case mix bias—Using opportunity for error rather than the number of patients treated as the denominator reduces the confounding that arises when one clinician or institution cares for sicker patients than another.¹⁴ This is because sicker patients present more opportunities for clinical process errors (of either omission or commission). Expressing errors as a function of opportunities for those errors adjusts (at least in part) for case

Comparison of effects of shifting and truncating the distribution of quality of care, assuming normal distribution

mix bias. This method cannot be used when outcomes are assessed because the patient is the smallest possible unit of aggregation under these circumstances.

Lack of stigma—The message is “improve X,” not “you are bad.” For this reason they are less likely to prompt perverse solutions. Arguably it is easier and more natural to improve the care process than to try to discredit the measure (see below).

Prompt wider action—Process measures encourage action from all organisations or individuals with room for improvement, not just a small proportion of outliers. Shifting the whole distribution will achieve a larger health gain than simply improving the performance of those in the bottom tail, as the figure shows. Assuming a normal distribution in quality, a shift of 10% would result in a health gain of 10%. However, improving the performance of the bottom 10% would produce a gain of 7.2%, even if this threshold distinguished perfectly between good and poorly performing units. Furthermore, organisations do not fail simultaneously across all dimensions of safety and quality. Rather they have particular strengths and weaknesses and improvement efforts can be targeted where they are needed: there is no need to produce a summary measure across criteria. In fact, we found no correlation between adherence to various evidence based quality criteria in 20 randomly selected UK maternity units.¹⁵ Thus, a hospital with above average recorded outcomes is still likely to have room for improvement in many aspects of care.

Useful for delayed events—Process measures are more useful than outcomes when the contingent adverse event is markedly delayed (such as failing to monitor patients with diabetes for proteinuria or to administer anti-D immunoglobulin when a rhesus negative woman gives birth to a rhesus positive baby).

Selecting and measuring clinical processes

Process standards used in performance management should be valid in that they must either be self evident measures of quality or be evidence based. However, validity is not sufficient—the standards must also be genuinely important to health care. This is because the opportunity cost of improving some processes may exceed the contingent gains.¹⁶ Worse, healthcare providers may put their efforts into the monitored processes at the expense of those that are not monitored.¹⁷ One way to ameliorate this effect may be to elicit clinical

standards from professional societies or consortia of providers and users of health care. There are plenty of important, evidence based criteria, and health services fall well short of full compliance.^{18 19} The Royal College of Obstetricians and Gynaecologists produced clinical guidelines as early as 1993²⁰ and a before and after study showed massive change in line with the evidence.¹⁵

The different methods of measuring process all have advantages and disadvantages. Broadly speaking, measures can be either explicit or implicit (although both methods can be combined). Explicit measures use a set of predetermined criteria (checklists). Implicit measures are assessed more like an expert review of a set of case notes. Although the explicit method has greater reliability (greater inter observer agreement), the implicit method covers more dimensions of care because it can detect errors that might not have been specified on a predetermined checklist.²¹

Is performance management effective?

Where outcomes are a specific measure of quality, externally imposed performance management by outcome may be effective. Collection of outcome data for cardiac surgery in the UK seems to have raised standards, although debate continues about whether the observed improvements exceed the secular trend.¹⁰ In the more common scenario where outcomes are not a specific measure of quality, process measures are a better method of judgment. However, process is expensive to measure as it currently requires access to patients' case notes. Evaluating case notes is time intensive and requires staff with clinical expertise. Electronic patients records and an increase in coded information in these records should make monitoring easier.

The cost of obtaining process measurements (and the contingent action) needs to be compared to the value (in terms of health benefit) of the improvement in quality that results from providers' responses to initial measurements. Although much of the evidence of effectiveness relates to bottom-up improvement programmes,²² there is also empirical support for the effectiveness of top-down performance management using process measures.^{23 24}

Are outcome measures obsolete?

Although process measures are the most suitable tool for performance management, measurement of outcomes remains important. Outcomes are useful for research, particularly for generating hypotheses. Here, simply finding an association between a variable (such as staff-patient ratios) and outcome (such as mortality) may be sufficient to prompt investigation, even when the strength of the association is low. Outcome data can also be used as a form of process control, such that institutions with abrupt changes in outcome or whose outcomes deviate by a large amount (three standard deviations or more is a sensible threshold²⁵) can be further investigated. For example, an outbreak of hospital acquired infection may be traced to a problem in the water supply. Lastly, the public are entitled to have access to outcome data, although such outcomes should always be published with a proper warning about the limitations.

Outcomes are useful for research, particularly for generating hypotheses

We thank the referees for helpful comments.

Contributors and sources: This paper is the result of a synthesis of prolonged writing, review, and consultation with experts in the field. RJL has had many opportunities to debate and test the argument presented here, most notably as a participant at numerous Pennyhill Anglo-American health summits. RJL conceived the article, which was subsequently drafted by RJL, CAB, and JN. All three authors approved the final manuscript. RJL is the guarantor.

Competing interests: None declared.

Provenance and peer review: Not commissioned, externally peer reviewed.

- 1 Esmail A. Failure to act on good intentions. *BMJ* 2005;330:1144-7.
- 2 Pitches D, Mohammed MA, Lilford RJ. What is the evidence that hospitals with higher risk-adjusted mortality rates provide poorer care? A systematic review of the literature. *BMC Health Serv Res* 2007;7:91.
- 3 Mant J, Hicks N. Detecting differences in quality of care: The sensitivity of measures of process and outcome in treating acute myocardial infarction. *BMJ* 1995;311:793-6.
- 4 Hayward RA, Hofer TP. Estimating hospital deaths due to medical errors: preventability is in the eye of the reviewer. *JAMA* 2001;286:415-20.
- 5 Brown CA, Lilford RJ. Cross sectional study of performance indicators for English PCTs: testing construct validity and identifying explanatory variables. *BMC Health Serv Res* 2006;6:81.
- 6 Lilford RJ, Mohammed M, Spiegelhalter D, Thomson R. Use and misuse of process and outcome data in managing performance of acute medical care: Avoiding institutional stigma. *Lancet* 2004;263:1147-57.
- 7 Deeks JJ, Dinnes J, D'Amico R, Sowden AJ, Sakarovich C, Song F, et al. Evaluating non-randomised intervention studies. *Health Technol Assess* 2003;7(27):1-173.
- 8 Nicholl J. Case-mix adjustment in non-randomised observational evaluations: the constant risk fallacy. *J Epidemiol Community Health* (in press).
- 9 Crow R, Gage H, Hampson S, Hart J, Kimber A, Storey L, et al. The measurement of satisfaction with healthcare: implications for practice from a systematic review of the literature. *Health Technol Assess* 2002;6(32):1-244.
- 10 Shortell SM, Jones RH, Rademaker AW, Gillies RR, Dranove DS, Hughes EF, et al. Assessing the impact of total quality management and organisational culture on multiple outcomes of care for coronary artery bypass surgery patients. *Med Care* 2000;38:207-17.
- 11 Bird SM, Cox D, Farewell VT, Goldstein H, Holt T, Smith OC. Performance indicators: good, bad and ugly. *J R Stat Soc A* 2005;168:1-27.
- 12 Locker TE, Mason SM. Are these emergency department performance data real? *Emerg Med J* 2006;23:558-9.
- 13 Parry GJ, Gould CR, McCabe CJ, O Tarnow-Mordi W. Annual league tables of mortality in neonatal intensive care units: longitudinal study. *BMJ* 1998;316:1931-5.
- 14 Lilford RJ, Mohammed M, Braunholtz D, Hofer TP. The measurement of active errors: Methodological issues. *Qual Saf Health Care* 2003;12(suppl 2):ii8-12.
- 15 Wilson B, Thornton JG, Hewison J, Lilford RJ, Watt I, Braunholtz D, et al. The Leeds University maternity audit project. *Int J Qual Health Care* 2002;14:175-81.
- 16 Hayward RA. Performance measurement in search of a path. *N Engl J Med* 2007;356:951-3.
- 17 Locker TE, Mason SM. Analysis of the distribution of time that patients spend in emergency departments. *BMJ* 2005;330:1188-9.
- 18 McGlynn EA, Asch SM, Adams J, Keesey J, Hicks J, DeCristofaro A, et al. The quality of health care delivered to adults in the United States. *N Engl J Med* 2003;348:2635-45.
- 19 Kirk SA, Campbell S, Kennell-Webb S, Reeves D, Roland MO, Marshall MN. Assessing the quality of care of multiple conditions in general practice. *Qual Saf Health Care* 2003;12:421-7.
- 20 Royal College of Obstetricians and Gynaecologists. *Effective procedures in obstetrics suitable for audit*. London: RCOG, 1993.
- 21 Lilford RJ, Edwards A, Girling A, Tanna GLD, Nicholl J, Hofer T. Inter-rater reliability measurements in the quality of health care: a systematic review. *J Health Serv Res Policy* 2007;12:172-80.
- 22 Jamtvedt G, Young JM, Kristoffersen DT, O'Brien MA, Oxman AD. Audit and feedback: effects on professional practice and health care outcomes. *Cochrane Database Syst Rev* 2006;(2):CD000259.
- 23 Lindenauer PK, Remus D, Roman S, Rothberg MB, Benjamin EM, Ma A, et al. Public reporting and pay for performance in hospital quality improvement. *N Engl J Med* 2007;356(5):486-96.
- 24 Kamerow D. Great health care, guaranteed. *BMJ* 2007;334:1086.
- 25 Mohammed MA, Cheng KK, Rouse A, Marshall T, Bristol, Shipman, and clinical governance: Shewhart's forgotten lessons. *Lancet* 2001;357:463-7.

SUMMARY POINTS

Process measures are the most suitable management tool for judging and rewarding quality
Clinical outcomes are likely to be affected by factors other than the quality of care
Outcome measures provide insufficient information about how to improve
Assessment of process encourages universal improvement rather than focusing on outliers
Selected measures must be valid and important