

Interpretation of hospital-specific outcome measures based on routine data

MICHAEL COORY, DANNY YOULDEN, AND PHILIP BAKER

Michael Coory is Medical Epidemiologist at the Health Information Centre, Queensland Health.

Danny Youlden is Senior Analyst in the Measured Quality Program, Queensland Health.

Philip Baker is Epidemiologist and Clinical Audit Information Manager, Office of the Chief Health Officer, Queensland Health.

Abstract

Hospital-specific outcome measures based on routine data are useful for stimulating interest in quality of care and for suggesting avenues for more in-depth analyses. They might also identify serious, once-in-a-lifetime failures of health care. However, such analyses are not definitive. They are a way of screening large amounts of routine data and, like all screening tools, they can generate false positives and false negatives. This is because differences in outcome measures across hospitals can be due to differences in types of patients seen (casemix), differences in data quality, and the play of chance; rather than differences in the quality of care. End-users of such analyses should be aware of these technical difficulties, otherwise skilled health workers in high-quality hospitals might be subjected to unwarranted criticism.

Background

In the United States, the production and dissemination of hospital report cards, based on routinely collected administrative data, is now a multimillion-dollar industry (Marshall et al., 2000). Similar analyses are done (or are being considered) in Canada, the United Kingdom and parts of Europe.

Australia is following this trend and in Queensland we have started a process of liaison with clinicians and administrators about the possible uses of such analyses. For cardiac surgery, analyses based on routinely collected administrative data have been used to augment more detailed audits based on retrospective chart review and patient surveys (Baker & Jenkins, 2002). Some preliminary experimental analyses have been also done for several medical indicators (Measured Quality Project, Queensland Health, 2002).

We have found that there is considerable confusion among clinicians and administrators about the purpose of outcome measures based on routine data. Further, the results of such analyses have been misinterpreted by the media (Sunday Program, Channel 9, 2002). The aims of this paper are to discuss the limitations of such analyses and to suggest how they fit into an overall strategy for improving quality of care.

Outcome measures from routine data

In Queensland, as in the rest of Australia and elsewhere in the developed world, routine hospital data contain only a limited number of variables that might be used as outcome indicators. These are in-hospital mortality, length-of-stay and readmission and these three outcomes have been extensively used in hospital-report cards in the United States and have formed the basis of many journal articles (Tu et al., 1999). Occasionally, depending on the condition of interest, other outcome measures can be created such as complication rates, but this is the exception rather than the norm.

Difficulties in interpretation

Experts agree that comparisons of outcomes (eg, mortality, length-of-stay, re-admission) across hospitals are difficult to interpret (Iezzoni, 1996). There are three main problems.

Differences in casemix (not comparing like with like)

Differences in outcome measures among hospitals may be due to differences in the types of patients seen (ie, differences in disease severity, coexisting conditions, age, smoking, nutrition, psychosocial factors, economic disadvantage, and the like). Statistical models can be used to adjust for these differences, but data might not be available for some potential confounders (eg, severity, smoking, nutrition, psychosocial factors, economic disadvantage) and for those where data are available, the quality of the information might be questionable. In any case, even with perfect data, statistical models can only reduce, not eliminate, the effects of casemix differences (Iezzoni, 1996).

Differences in data quality

Differences among hospitals in the measurement of the adverse outcome of interest (eg, readmission) or the variables used in the casemix adjustment will falsely lead to apparent differences in the outcome indicator. For example, an important factor in explaining an observed decrease in casemix-adjusted mortality from cardiac surgery in New York State was an apparent increase in the prevalence of risk factors in the patients who had surgery. Between 1989 and 1991, there was an increase in the reported prevalence of renal failure, congestive heart failure, chronic obstructive pulmonary disease and unstable angina in the patients who had surgery. These increases were due to increased reporting, rather than genuine changes in casemix and accounted for 40% of the observed decrease in casemix-adjusted mortality (Green & Winfield, 1995).

Chance

Even if the casemix adjustment and data quality could be perfect (and it cannot), outcome measures for hospitals would still be difficult to interpret because they are vulnerable to the play of chance. For example, Poloniecki and his co-workers (1998) found that year-to-year differences in mortality following heart surgery were large (odds ratio 1.5) even when the underlying mortality and casemix did not change. Similarly, using data from fertility clinics, Marshall and Spiegelhalter (1998) showed that there was great uncertainty about the rankings based on livebirth rates. Many centres had substantial changes in ranks between years, even though their live birth rate did not change significantly.

We have analysed data (casemix-adjusted) on mortality following acute myocardial infarction for hospitals in Queensland for a recent five-year period. We found considerable year-to-year variation. For example one hospital out of 28 went from the 2nd worst in 1995/96 to the 9th worst in 1996/97 to the 23rd worst (5th best) in 1997/98 to the 4th worst in 1998/99 and then the 22nd worst (6th best) in 1999/00.

Figure 1: reasons for differences in hospital-specific outcome measures

•	Differences in casemix
•	Differences in data quality
•	Chance
•	Differences in quality of care

In short, besides differences in the quality-of-care, three possibilities need to be considered when differences are observed among hospitals for a particular outcome measure (Figure 1). Because of these three possible explanations and because no statistical method can completely account for them, interpretation of hospital-specific outcome measures based on routine data can be problematic.

A framework for using hospital-specific outcome measures to improve quality of care

Because of the above caveats, analyses of outcome measures cannot be definitive. They are best viewed as a screening tool to stimulate interest in quality at the hospital, regional and state level, and to suggest useful avenues for further investigation. This approach is attractive because in-depth evaluations are costly and there is a need to identify where to target scarce resources for improving quality of care (Iezzoni, 1996).

A report by Tu et al (1999) from Ontario points out that ‘... *screening tests such as Pap smears or mammograms are often used in medicine and these screening tests produce both false positives (women with a positive test who do not have cancer) and false negatives (women with cancer who have a negative test). Screening tests can help to identify cases that need follow up. The same is true for measures of comparative hospital performance. An effort is made to minimise false positives, but they cannot be eliminated. Thus, the measures of clinical performance [from routine hospital data] ... should be taken not as a definitive assessment of the quality of care, but rather the first step in a process of quality improvement that should involve more detailed analysis at every institution.*

Figure 2: role of hospital-specific outcome measures based on routine data

-
- Screen routine data to stimulate interest in quality and suggest avenues for more in-depth analyses.
 - Identify serious, ‘once-in-a-lifetime’ failures of health care (eg, the Bristol case).
-

Rarely, serious failures of health care may occur and, in these situations, analyses of routine data can serve as an early warning system. One example from the United Kingdom is the Bristol case in which three doctors were found guilty by the General Medical Council of *serious professional misconduct* in relation to the deaths of 29 babies and young children (Bolsin, 2001). A recent analysis suggests that monitoring outcomes through routine data could have identified the significant deviations from expected mortality rates (Mohammed *et al.*, 2001).

Process measurement: the way of the future?

Process measurement is an assessment of the degree to which health care adheres to processes that are proven by scientific evidence to affect health. Examples include the proportion of eligible patients with acute myocardial infarction who receive thrombolysis, or the proportion of such patients who receive coronary rehabilitation on discharge.

Process measures are attractive because once the eligible population is defined, casemix adjustment is generally not necessary. Further, process measures provide information that is actionable; it tells clinicians what is being done well and what needs improvement (Rubin *et al.*, 2001).

In general, routine hospital databases do not contain data for process measurement. It is hoped that hospitals would respond to outcome measurement based on routine data by developing and implementing their own evidence-based process measures, perhaps through the use of clinical pathways (Antioch *et al.*, 2001). In the future, improvements in information technology might make it possible to collect process measures routinely and on a statewide or national basis via high-quality clinical databases (Black, 1999).

Comment

End-users of outcome measures based on routine data need to be aware of their limitations and how they fit into the totality of efforts aimed at improving quality of care. Such analyses are now conducted in several countries, not because they provide definitive, error-free measures of quality of care, but because they are cost-effective. The routine data must be collected anyway (under the terms of the Health Care Agreement between the States and Territories and the Commonwealth) and therefore the outcome measures can be generated at small marginal cost. We believe that the limitations of such analyses (problems with casemix, data quality and

accounting for the play of chance) must be made explicit to all users (eg, clinicians, administrators, the public, the media) otherwise skilled health workers in high-quality hospitals will be subjected to unwarranted criticism.

References

- Antioch K, Chapman R, Elliott B, Santamaria N, Crawford R, Fiddes K 2001, 'Cost-effective clinical pathways at The Alfred Hospital: international lessons from Bayside Health, Australia', *Australian Health Review*, vol 24, pp21-29.
- Baker P & Jenkins S 2002, 'Using routinely collected hospital data for risk-adjusted analyses to permit benchmarking for quality improvement', [abstract] *Proceedings of Second Asia Pacific Forum on Quality Improvement in Health Care*, [in press].
- Black N 1999, 'High quality clinical databases: breaking down barriers', *BMJ*, vol 353, pp 1205-1206.
- Bolsin S 2001, 'League tables and professionalism: on quality in health care', *Australian Health Review*, vol 24, pp1-4.
- Green J & Winfield N 1995, 'Report cards on cardiac surgeons: Assessing New York State's approach', *N Engl J Med*, vol 332, pp 1229-1232.
- Iezzoni LI 1996, 'Assessing quality using administrative data', *Ann Int Med*, vol 127, pp 666-674.
- Marshall EC & Spiegelhalter DJ 1998, 'Reliability of league tables of in vitro fertilisation clinics: retrospective analysis of live birth rates', *BMJ*, vol 316, pp 1701-1705.
- Marshall MN, Shekelle PG, Leatherman S, Brook RH 2000, 'The public release of performance data. What do we expect to gain? A review of the evidence', *JAMA*, vol 283, pp 1866-1874.
- Measured Quality Project, Queensland Health 2002, 'Measured Quality', available at: <http://www.health.qld.gov.au/quality/> (accessed 29/05/02).
- Mohammed MA, Cheng KK, Rouse A, Marshall T 2001, 'Bristol, Shipman, and clinical governance: Shewhart's forgotten lessons', *Lancet*, vol 357, pp 463-467.
- Poloniecki J, Valencia O, Littlejohns P 1998, 'Cumulative risk adjusted mortality chart for detecting changes in death rate: observational study of heart surgery', *BMJ*, vol 316, pp 1697-1700.
- Rubin HA, Pronovost P, Diette B 2001, 'The advantages and disadvantages of process-based measures of health care quality', *Int J Qual Health Care*, vol 13, pp 469-474.
- Sanderson C & McKee M 1998, 'How robust are rankings? The implications of confidence intervals', *BMJ*, vol 316, pp 1705.
- Smith R 1998, 'All changed, changed utterly', *BMJ*, vol 316, pp 1917-1918.
- Sunday Program, Channel Nine 2002, 'A hospital's secret', 14 April 2002, available at: http://news.ninemsn.com.au/Sunday/cover_stories/ (accessed 29/05/2002).
- Tu JV, Austin P, Naylor CD, Iron K, Zhang H 1999, 'Acute myocardial infarction outcomes in Ontario', in CD Naylor, PM Slaughter (eds) *Cardiovascular health and services in Ontario: an ICES atlas*, Institute for Clinical Evaluative Sciences, Toronto, pp 83-110.