

Are diagnosis specific outcome indicators based on administrative data useful in assessing quality of hospital care?

I Scott, D Youlden, M Coory

Qual Saf Health Care 2004;13:32–39. doi: 10.1136/qshc.2002.003996

Background: Hospital performance reports based on administrative data should distinguish differences in quality of care between hospitals from case mix related variation and random error effects. A study was undertaken to determine which of 12 diagnosis-outcome indicators measured across all hospitals in one state had significant risk adjusted systematic (or special cause) variation (SV) suggesting differences in quality of care. For those that did, we determined whether SV persists within hospital peer groups, whether indicator results correlate at the individual hospital level, and how many adverse outcomes would be avoided if all hospitals achieved indicator values equal to the best performing 20% of hospitals.

Methods: All patients admitted during a 12 month period to 180 acute care hospitals in Queensland, Australia with heart failure (n=5745), acute myocardial infarction (AMI) (n=3427), or stroke (n=2955) were entered into the study. Outcomes comprised in-hospital deaths, long hospital stays, and 30 day readmissions. Regression models produced standardised, risk adjusted diagnosis specific outcome event ratios for each hospital. Systematic and random variation in ratio distributions for each indicator were then apportioned using hierarchical statistical models.

Results: Only five of 12 (42%) diagnosis-outcome indicators showed significant SV across all hospitals (long stays and same diagnosis readmissions for heart failure; in-hospital deaths and same diagnosis readmissions for AMI; and in-hospital deaths for stroke). Significant SV was only seen for two indicators within hospital peer groups (same diagnosis readmissions for heart failure in tertiary hospitals and inhospital mortality for AMI in community hospitals). Only two pairs of indicators showed significant correlation. If all hospitals emulated the best performers, at least 20% of AMI and stroke deaths, heart failure long stays, and heart failure and AMI readmissions could be avoided.

Conclusions: Diagnosis-outcome indicators based on administrative data require validation as markers of significant risk adjusted SV. Validated indicators allow quantification of realisable outcome benefits if all hospitals achieved best performer levels. The overall level of quality of care within single institutions cannot be inferred from the results of one or a few indicators.

See end of article for authors' affiliations

Correspondence to:
Dr I Scott, Director of
Internal Medicine, Princess
Alexandra Hospital,
Ipswich Road, Brisbane,
Queensland, Australia.
4102; ian_scott@
health.qld.gov.au

Accepted 6 July 2003

Routinely collected administrative data are increasingly being used to evaluate the quality of care provided by individual hospitals.¹ Profiling of hospital performance with regard to mortality, length of stay, and readmissions for various cardiovascular diseases is now commonplace in Canada,² the United States,^{3–5} and the UK.⁶ In Australia, considerable work has been undertaken by the Australian Council of Health Care Standards in developing and measuring sets of clinical indicators based on hospital inpatient data which reveal whether variations exist between hospitals that might suggest potential for improving the quality of care.^{7–8}

However, no randomised trial has evaluated the effects of regular reporting of hospital specific performance indicators on quality of care or patient outcomes.⁹ Moreover, report cards have been described as untimely, lacking in process of care information, subject to confounding bias, devoid of robust comparisons between hospitals of similar type (peer hospitals), and exerting only modest effects within quality improvement programmes.^{10–11}

Various methodological limitations of reported indicators have been enunciated: (1) absence of consistent correlation between indicators based on administrative data and chart based process of care audits^{12–13}; (2) potential for bias as a result of inadequate adjustment for differences between hospitals in case mix and illness severity^{14–15}; (3) lack of data

quality verification¹⁶ and standardisation of data collection methods¹⁷; (4) limited generalisability of ratings of hospital performance based on single or even multiple indicators^{18–19}; and (5) risk of labelling hospitals with smaller caseloads as being poor performers on the basis of results which may occur simply by chance.^{20–21}

Our study aimed to identify diagnosis-outcome indicators which constitute measures of systematic variation in performance which reflect potential quality of care problems that may warrant more detailed investigation in individual hospitals. Any observed variation between hospitals in diagnosis-outcome indicators (such as in-hospital mortality for patients with stroke) may be explained in part by two basic types of variation: (1) common cause variation due to measurement error, differences in patient characteristics (or case mix), or simply the play of chance (particularly in the case of small numbers of admissions); and (2) special cause or systematic variation due to real differences between hospitals in their quality of care.²² If the reported results of any set of clinical indicators are to be interpreted correctly in making impartial judgements about quality of care at a particular hospital, one must be confident that these indicators have been statistically validated as markers of real or systematic variation in care.

Health professionals within hospitals may also be interested in comparing the performance of their hospital with

that of other hospitals which bear a close resemblance in terms of referral status and level of services provided.²³ Accordingly, it would be helpful to determine the extent to which systematic variation in indicator results, seen in analyses which compare all hospitals, persists when analysis is restricted to subsets of peer group hospitals comprising only tertiary, community, or district hospitals.

We were also interested to determine whether results for different diagnosis-outcome indicators were correlated at the level of individual hospitals—that is, whether performance for one indicator predicted a similar performance for others. Care of suboptimal quality may be a generic trait within a particular institution regardless of the clinical condition being managed, or it may be specific to one or a select number of patient groups defined by the principal diagnosis.^{24–25}

If we can identify indicators which can reliably detect below average performers, this may enable more precise quantification of the number of adverse outcome events (death, readmissions, or long stays) that might be avoided if all hospitals under study were to achieve the same results as those of above average performers.²⁶

In addressing these issues, we undertook a study of administrative data from 180 public hospitals in the state of Queensland, Australia relating to patients admitted with either one of three cardiovascular diseases: heart failure, acute myocardial infarction (AMI), or stroke. These conditions were chosen for study as they are common causes of acute hospitalisation, are associated with a significant burden of morbidity and mortality, and for which optimal care has been well defined by the results of numerous clinical trials. These conditions also feature prominently in other studies of clinical performance and quality improvement.^{2–6}

For each clinical condition (or diagnosis) we evaluated four key outcomes: in-hospital mortality; length of hospital stay; same cause readmission at 30 days; and all cause readmission at 30 days. Each coupling of a specific diagnosis with a specific outcome measure constituted a distinct “diagnosis-outcome indicator”.

The objectives of the study were as follows:

- To measure the degree of systematic variation between hospitals for each diagnosis-outcome indicator after minimising variation due to differences in case mix or secondary to random error using appropriate statistical methods.
- For those diagnosis-outcome indicators identified as being candidate indicators of quality (that is, showing significant systematic variation across all hospitals), to determine:
 - whether significant systematic variation seen at the individual hospital level persisted at the group level when hospitals were grouped into tertiary, community, or district hospitals;
 - the extent to which, for individual hospitals, poor performance for one diagnosis-outcome indicator predicted similar findings for other diagnosis-outcome indicators;
 - potential aggregate savings in adverse outcomes if all hospitals were to achieve indicator values similar to those of the best performing 20% of hospitals.

METHODS

Data sources

In the state of Queensland, Australia, approximately 1 100 000 episodes of care occurred during the financial year 1999/2000 in 180 acute care hospitals serving a population of

3.5 million residents. Most hospitals are low volume institutions; only 34 had more than 10 000 episodes of care and a further 21 had 5000–10 000 episodes of care. These 55 hospitals accounted for 81% of all hospital episodes of care. A description of the Australian hospital system is shown in box 1.

Data relating to all episodes of acute care for the financial year 1999/2000 were obtained from the Queensland Hospitals Admitted Patients Data Collection (QHAPDC, box 2).²⁷ The coding method used during the study period was the International Classification of Diseases and Health Related Problems, 10th revision, Australian Modification (ICD-10-AM).²⁸ Up to 10 separate diagnoses can be listed for each abstract, with one nominated as the principal diagnosis chiefly responsible for the patient’s admission to hospital. Records from QHAPDC were included on the basis of the principal diagnosis code only.

Study subjects

The chosen diagnostic codes were heart failure (ICD-10-AM code I50), acute myocardial infarction (AMI) (I21, I22), and stroke (I61–I64). To reduce variation in diagnostic and coding accuracy, the following exclusion criteria were applied to all QHAPDC records for each diagnosis based on published literature²⁹ and local expert opinion: (1) age <30 or >89 years; (2) non-acute (or elective) patients; (3) inter-hospital transfers; (4) length of acute hospital stay (LOS) >30 days; and (5) usual residence either interstate or overseas. In addition, in the case of AMI and stroke, admissions with LOS of <4 days and <3 days, respectively, but discharged alive were excluded in order to minimise cases of misdiagnosis. In patients with stroke, those in which elective carotid endarterectomy was performed during the same admission were also excluded.

Outcome events

30 day in-hospital mortality

Potentially avoidable deaths secondary to suboptimal care may manifest as premature deaths in patients with a low or moderate baseline risk—that is, in patients who are expected to survive due to less severe illness or comorbidity. In-hospital death occurring before 30 days was the definition chosen on the assumption that deaths after 30 days were

Box 1 Australian hospital system

In Australia the healthcare system comprises two hospital sectors:

- Public sector in which hospital care, both inpatient and outpatient, is funded free of charge by state governments with triennial block funding provided by the federal government under Medicare health agreements.
- Private sector where hospitals and the doctors who provide services within them charge fees to patients which comprise two portions: a rebatable portion paid for by the federal government (equal to 75–85% of the Medicare Benefits Schedule fee) and the remainder which patients may elect to pay fully themselves as an out of pocket expense or for which they seek a refund for all or some of the amount from a health insurance fund to which they subscribe by paying an insurance premium. At the present time this premium is also subsidised by the federal government by a 30% taxation rebate.

Box 2 Administrative database

In Queensland, all hospitals—both public and private—are required for every episode of in-hospital care to submit a discharge abstract to a centrally located administrative database maintained by the state health department. An episode of care ends when the principal intent changes or when the patient is formally discharged from the facility. The discharge abstract details patient demographic data, coded principal diagnosis or procedure, complications and comorbidities, and admission and discharge dates. These abstracts comprise what is termed the Queensland Hospitals Admitted Patients Data Collection (QHAPDC).²⁷

more likely to be secondary to advanced illness, serious comorbid conditions, or other factors not associated with quality of care.³⁰ Fewer than 5% of in-hospital deaths occurred after 30 days. Deaths which occurred out of hospital were excluded as these were not identifiable from existing databases. Patients who died in hospital were excluded from analyses of all other outcome events.

Long stays

A long stay was defined as an episode of acute care which exceeded the 90th percentile of all length of stays for the diagnosis of interest. In calculating this indicator, patients with AMI who had an invasive coronary procedure during the same episode of care were excluded. Length of stay was calculated on the basis of days of hospitalisation for which episode of care classification was deemed to be acute care—that is, days of hospitalisation comprising non-acute care such as rehabilitation or awaiting residential care placement were excluded.

30 day readmissions

These were defined as acute (non-elective) patient readmissions to the same hospital within 30 days of discharge. Readmissions to other hospitals could not be traced as the unit record identifier on the QHAPDC is unique to each hospital. Readmissions were of two types: all cause readmissions and readmissions with the same principal diagnosis as the index admission.

Risk factors

We adjusted outcome data for age, sex, illness complications and comorbidities, assuming that any residual differences in outcome reflected differences in quality of care received. The following comorbidities relevant to cardiovascular disease were included in adjustment models:³¹ malignancy, diabetes with and without complications, dementia, valvular disorders, hypertension, ischaemic heart disease, cardiomyopathy, conduction disorders, dysrhythmias, heart failure, stroke, peripheral vascular disease, acute lower respiratory tract infection, chronic obstructive pulmonary disease, liver disease, and renal failure. Complications used in adjusting results included hypotension and shock, hyponatraemia, and anaemia.

Hospital groupings

In determining whether systematic variation in indicator results at the level of all hospitals persisted when the analysis was conducted at the level of peer hospital groupings, we defined a subset of hospitals comprising the 57 largest public hospitals in the state (based on annual admission numbers) and grouped these according to teaching and referral status into one of three peer groups: (1) tertiary public hospitals (n = 4); (2) community public hospitals (n = 16); and (3)

district public hospitals with an annual budget in excess of \$AUS2 million (n = 37).

Statistical analysis

At the level of the individual patient records, diagnosis specific outcome data were adjusted for identified risk factors and hierarchical statistical methods were then used to detect systematic variation in outcomes after records were aggregated at the level of individual hospitals. Hierarchical (or multilevel) statistical methods involve specifying two or more levels (or stages) of relationships among study variables—for example, at the level of patients, hospitals, or the whole state. This approach involves partitioning the variance into components attributable to individual factors (such as within hospital variation) and that which is attributable to higher level factors (such as between hospital variation).³² The sequential steps undertaken were as follows. All analyses were conducted using SAS statistical software package.³³

Identifying candidate risk factors for risk adjustment models

For each diagnosis-outcome indicator, bivariate analysis was undertaken to show the strength of the association between the risk factors previously defined and the outcome event of interest across all hospitals. Those risk factors whose crude outcome odds ratios had p values <0.25 were included in risk adjustment models (see below).³⁴

Calculating risk adjusted ratios of observed to expected outcome event rates for each diagnosis for each hospital

For each diagnosis-outcome indicator, logistic regression models were applied to all observations (all hospitals combined) using the PROC LOGISTIC procedure in SAS in calculating expected outcome event rates for each hospital. These models were adjusted for age (in 5 year categories), sex, and risk factors in calculating an estimated probability that the episode of care would result in the outcome being investigated. The estimates for all discharges from a particular hospital were summed to obtain the expected values which were then compared with the observed values as observed/expected (O/E) ratios. Model discrimination was assessed using the “c” statistic, which is equivalent to the area under the receiver-operating characteristic (ROC) curve and is a measure of the rank correlation between the observed outcome and the predicted probability at the level of the individual observations.³⁵ Model calibration was determined using a goodness of fit test.³⁴

Apportioning variation in O/E ratios to random v systematic variation

We partitioned the variation in the O/E ratios across hospitals into that due to chance (more likely in low volume hospitals) and that due to systematic variation.³⁶ A hierarchical model was used to estimate: (1) random variation of the observed O/E ratios around the true O/E ratios within each hospital; and (2) variation in the true O/E ratios (systematic variation) across hospitals. The maximum likelihood value (with 95% confidence interval) for the systematic variation was calculated using a method described by Martuzzi and Hills.³⁷ Systematic variation was considered to be statistically significant if the 95% confidence interval excluded zero to three decimal places. This analysis allowed us to determine whether all, or only some, of the diagnosis-outcome indicators were capable of distinguishing between hospitals on the basis of real differences in quality of care, as validated statistically by rejecting the null hypothesis that no systematic variation existed.

Determining systematic variation in indicators at the level of peer hospital groupings

For those diagnosis-outcome indicators associated with significant systematic variation for all hospitals, the process outlined above was repeated within the three peer hospital groups previously defined. We wished to determine whether systematic variation at the level of all hospitals persisted or was extinguished if the analyses were repeated at the level of peer hospital groups—that is, whether the ability to discriminate between low and high quality care hospitals was lost when hospitals with similar characteristics were compared.

Correlation analyses

For those diagnosis-outcome indicators associated with significant systematic variation at the all-hospital level, correlations between different indicators were determined on the basis of their individual outcome O/E ratios (Pearson coefficients, r) or as ratios grouped as quintiles (Spearman rank coefficients, ρ). This analysis was undertaken to assess the extent to which, at the individual hospital level, different diagnosis-outcome indicators gave similar results.

Quantification of potential savings in outcome events

Finally, the number of outcome events that could potentially be avoided at the whole state level was calculated for those diagnosis-outcome indicators associated with systematic variation based on all hospitals achieving an outcome O/E ratio equal to the 20th percentile of the O/E ratio distribution.

RESULTS

Cohort characteristics and outcome measures

Patient characteristics for each diagnosis are shown in table 1. In-hospital mortality was highest for stroke (20.8%) and lowest for heart failure (8.0%). Patients with stroke also had the longest stay in hospital (median 8 days), while 30 day all cause readmission rates to the same hospital ranged from 12.0% for stroke to 28.1% for heart failure.

Risk adjustment regression models

The risk factors included in risk adjustment regression models are shown in table 2. After applying these models to each diagnosis-outcome indicator (table 3), discrimination was highest (“ c ” statistic >0.7) for 30 day in-hospital mortality for heart failure and AMI and long stays for AMI. All models were well calibrated (goodness of fit statistic p values >0.05) except for in-hospital mortality for AMI, where further partitioning suggested this was due to the relatively large number of observations rather than the model providing a poor fit.

Estimations of systematic variation

All hospitals

Comparing all hospitals, systematic variation was found to be significant for long stays (0.176, 95% CI 0.071 to 0.354) and same diagnosis readmissions (0.068, 95% CI 0.013 to 0.172) for heart failure; for in-hospital mortality (0.122, 95% CI 0.049 to 0.260) and same diagnosis readmissions (0.135, 95% CI 0.039 to 0.347) for AMI; and for in-hospital mortality (0.048, 95% CI 0.004 to 0.132) for stroke (table 4). For indicators associated with significant systematic variation, risk adjusted O/E event ratios differed by as little as 1.6-fold (adjusted O/E ratio 0.81–1.33) for same diagnosis readmissions due to AMI up to almost sixfold (adjusted O/E ratio 0.35–2.00) for long stays due to heart failure (table 4).

Peer group hospitals

Comparing peer group hospitals (table 5), systematic variation remained significant for only two indicators within two hospital peer groups: same diagnosis readmissions for heart failure in tertiary hospitals (0.115, 95% CI 0.009 to 0.848) and in-hospital mortality for AMI in community hospitals (0.103, 95% CI 0.020 to 0.323).

Indicator correlations

For those indicators associated with significant all hospital systematic variation, only one pair showed any significant correlation based on individual O/E ratios: a relatively weak negative relation between stroke mortality and long stays for heart failure ($r = -0.179$; $p = 0.044$). Two pairs of indicators were correlated when ratios were grouped as quintiles: same diagnosis readmission for heart failure and stroke mortality ($\rho = 0.213$; $p = 0.017$) and AMI and stroke mortality ($\rho = 0.409$; $p = 0.008$).

Potentially achievable reductions in outcome events

Selecting only indicators associated with significant systematic variation, nearly 30% of in-hospital deaths due to AMI (132 deaths per year) and almost 20% of deaths due to stroke (115 deaths per year) could be avoided if all hospitals in Queensland were to achieve the same O/E mortality ratios as the top performing 20% of hospitals in the state (table 6). Top level performance would similarly result in avoidance of over one third of long stays due to heart failure, more than one fifth of same diagnosis readmissions for heart failure, and almost one third of same diagnosis readmissions for AMI.

DISCUSSION

Main findings

Statistically significant systematic variation between hospitals in diagnosis-outcome indicators may not be entirely explained by differences in quality of care. A portion of the level of variation may still be due to data quality or reporting

Table 1 Cohort characteristics and summary outcome measures

Characteristic	Selected medical diagnoses		
	Heart failure	AMI	Stroke
Total cohort (no of discharges)	5745	3427	2955
Women (%)	47.9	36.8	48.4
Mean (SD) age (years)	75.1 (10.3)	68.8 (12.8)	74.3 (10.9)
30 day in-hospital mortality rate (%)	8.0	12.9	20.8
Length of stay			
Median (days)	5	6	8
90th percentile (days)	14	12	19
30 day readmission rate*			
All causes (%)	28.1	17.9	12.0
Same principal diagnosis (%)	11.4	12.8	3.1

*Readmissions were unplanned to the same hospital; readmissions for the same principal diagnosis for acute myocardial infarction (AMI) included all ischaemic heart disease (ICD10 codes I20–I25).

Table 2 Test of association between risk factors and diagnosis specific outcomes

Risk factor (comorbidity/complication)	Diagnosis-outcome indicator											
	Heart failure				AMI				Stroke			
	M	L	RA	RS	M	L	RA	RS	M	L	RA	RS
Malignancy	?	?	?		?		?		?		?	?
Diabetes without complications		?		?			?					
Diabetes with complications		?	?	?		?	?	?	?		?	?
Hyponatraemia	?	?		?	NA	NA	NA	NA		?	?	
Anaemia		?	?		NA	NA	NA	NA		?		
Dementia	?	?		?			?		?	?	?	?
Valvular disorders			?		NA	NA	NA	NA	?			
Hypertension	?	?		?	?		?		?	?	?	
Ischaemic heart disease	?	?	?		NA	NA	NA	NA	?		?	?
Cardiomyopathy	?			?	NA	NA	NA	NA			?	
Conduction disorders					?	?		?	?	?		
Dysrhythmias	?	?			?	?		?	?	?		
Heart failure	NA	NA	NA	NA	?	?	?		?	?	?	
Stroke	?	?	?	?	?	?			NA	NA	NA	NA
Peripheral vascular disease			?		?		?					
Acute LRTI and influenza	?	?		?	?	?	?		?	?		
COPD	?		?		?	?	?		?	?	?	?
Liver disease	?	?					?		?			
Renal failure	?	?	?		?	?	?		?	?	?	

? indicates significant association on bivariate analysis ($p < 0.25$ for corresponding odds ratio); blank cells indicate $p \geq 0.25$; NA = not applicable (risk factor was not considered in the analysis for that diagnosis-specific outcome); M = 30 day in-hospital mortality; L = length of stay exceeding 90th percentile; RA = 30 day readmissions to the same hospital for all causes; RS = 30 day readmission to the same hospital for the same principal diagnosis; AMI = acute myocardial infarction; LRTI = lower respiratory tract infections; COPD = chronic obstructive pulmonary disease.

differences between hospitals or inadequate case mix adjustment. However, we have attempted to minimise the effects of the former by using a standardised central administrative database that applies to all hospitals, and the effects of the latter by employing appropriate risk adjustment and hierarchical statistical methods. We contend that, as a result of such analysis, those indicators associated with systematic variation can be validly used as a tool to screen for potentially suboptimal care across hospitals.

In-hospital mortality for AMI and stroke, same diagnosis readmissions for AMI and heart failure, and long hospital stays for heart failure were the indicators showing significant systematic variation. These results suggest that not all diagnosis-outcome indicators can be used as reliable markers of interhospital variation in quality of care. Of 12 diagnosis-outcome indicators relating to three cardiovascular diagnoses, only five (42%) were associated with significant systematic variation across all hospitals. Of these, only two displayed significant variation when peer group hospitals were compared, and this was seen within only two hospital peer groups.

In terms of rating overall hospital performance, only four diagnosis-outcome indicators (33%) showed any degree of correlation at the level of individual hospitals. The inverse relation between stroke mortality and long stays for heart

failure seems implausible and, given its marginal statistical significance within multiple comparisons, we would question the meaningfulness of this relation. On more solid ground, based on comparisons of results expressed in quintiles, higher mortality rates for stroke appear to predict higher mortality rates for AMI and higher same diagnosis readmission rates for heart failure.

With regard to quantification of potentially avoidable deaths due to suboptimal quality of care, statistically significant systematic variation between all hospitals was seen for AMI and stroke mortality, and between peer grouped community hospitals for AMI mortality. This variation was manifested as 2–3-fold differences in standardised mortality ratios for all hospitals which, if shifted to best performer values (20th percentile), would result in a combined saving in Queensland of more than 200 lives annually.

Study limitations

Our study has several limitations. The first relates to sample size for each indicator in that larger numbers of admissions to individual hospitals may have resulted in more indicators having significant variation. However, within many health jurisdictions throughout developed countries²⁻⁵ the distribution of hospital admission volumes resembles that seen here, and most report cards present analyses of data collected at a

Table 3 Test statistics for risk adjustment regression models

Diagnosis	Outcome			
	In-hospital mortality	Long stays	Readmissions all causes	Readmissions same principal diagnosis
Heart failure	c = 0.719 p = 0.146	c = 0.683 p = 0.125	c = 0.601 p = 0.824	c = 0.617 p = 0.289
AMI	c = 0.793 p = 0.006	c = 0.808 p = 0.220	c = 0.645 p = 0.371	c = 0.567 p = 0.440
Stroke	c = 0.682 p = 0.075	c = 0.654 p = 0.393	c = 0.636 p = 0.817	c = 0.627 p = 0.759

AMI = acute myocardial infarction.
c = "c" statistic corresponding to the area under the receiver operating characteristics (ROC) curve. Statistic values >0.7 indicate the model discriminates well; value of 0.5 indicates discrimination no better than chance.
A p value of >0.05 indicates the model provides a good fit for the data (Hosmer-Lemeshow goodness of fit test).

Table 4 Systematic variation (SV) and range of risk adjusted observed/expected (O/E) event ratios for diagnosis-outcome indicators*

Diagnosis	Outcome			
	30 day in-hospital mortality	Long stays	30 day all cause readmissions	30 day same diagnosis readmissions
Heart failure				
SV	0.067	0.176*	0.008	0.068*
95% CI	0.000 to 0.204	0.071 to 0.354	0.000 to 0.031	0.013 to 0.172
O/E ratio	0.89-1.22	0.35-2.00	0.81-1.18	0.54-1.87
AMI				
SV	0.122*	0.114	0.002	0.135*
95% CI	0.049 to 0.260	0.000 to 0.384	0.000 to 0.051	0.039 to 0.347
O/E ratio	0.63-1.99	0.79-1.13	0.95-1.05	0.81-1.33
Stroke				
SV	0.048*	0.012	0.031	0.122
95% CI	0.004 to 0.132	0.000 to 0.130	0.000 to 0.147	0.000 to 0.626
O/E ratio	0.74-1.58	0.64-1.34	0.83-1.27	NA

*Results for all hospital analysis. Systematic variation (SV) was regarded as significant if the 95% confidence interval (CI) excluded zero to three decimal places. Range of risk adjusted O/E event ratios apply only to those hospitals where expected number of outcome events for 1 year was ≥5. AMI = acute myocardial infarction; NA = not applicable as no individual facility had outcome events which numbered ≥5.

state or provincial level. Some authors have suggested it would be better, in the interests of quality improvement, to concentrate on the point estimate of the systematic variation and to disregard non-significant confidence intervals.³⁶ However, hospitals may not wish to invest limited resources in improving quality of care of a particular patient group if the observed differences from state averages in their indicator results might be due to chance alone.

Secondly, we only examined data for three conditions over a 12 month period. Whether indicator values for individual hospitals show sustained trends over time is difficult to ascertain given considerable year to year fluctuations and a paucity of robust statistical methods for smoothing such fluctuations.³⁸ However, the rapid changes in clinical practice and demand for timely indicator reporting mean that extending the observation period into the more remote past reduces clinical relevance.

Thirdly, we did not attempt to correlate, and therefore externally validate, indicators associated with systematic variation (based on administrative data) with explicit process of care audits of individual patients. The literature, which has mainly concentrated on mortality, yields conflicting opinions on this issue. Keeler *et al* concluded that, even with rigorous risk adjustment, standardised mortality ratios show weak correlations with process of care measures.³⁹ In contrast, Kahn and colleagues found that standardised mortality was higher in patients receiving poorer quality of care, as measured by explicit criteria, for medical diagnoses of heart failure, AMI, stroke, and pneumonia.¹²

Implications for practice

On the basis of our results we would caution against the indiscriminate use of indicators based on administrative data to flag potential variations in quality of care in the absence of appropriate statistical confirmation that such indicators show systematic variation in their results. We would recommend that any diagnosis-outcome indicator should be shown by such methods to measure differences between hospitals that are not simply due to differences in case mix or random error due to small numbers of admissions. In this study, in-hospital mortality for AMI and stroke, same diagnosis readmissions for heart failure and AMI, and length of stay for heart failure appear to be discriminatory indicators across all hospitals. At the community hospital level, in-hospital mortality for AMI continues to be discriminatory, as is same diagnosis readmission for heart failure at tertiary hospital level.

The paucity of correlation between values of different diagnosis-outcome indicators in individual hospitals also challenges the validity of conclusions about the overall quality of individual hospital care for patients with cardiovascular disease based on results of a single or a selected number of indicators. Chassin *et al* found little correlation between rankings of hospital mortality rates for 22 diagnoses adjusted for age, race, and sex.²⁴ Rosenthal and colleagues found similar results for seven high volume diagnoses among hospitals with large sample sizes,^{19, 25} even when data were aggregated over several years. Using Bayesian modelling of mortality estimates for different conditions based on

Table 5 Systematic variation (SV) for diagnosis-outcome indicators within peer group hospitals*

Diagnosis-outcome indicator	Hospital type		
	Tertiary	Community	District
Heart failure			
Long stays (14-30 days)	0.020 (0.000 to 0.509)	0.097 (0.000 to 0.432)	0.007 (0.000 to 0.653)
30 day same diagnosis readmissions	0.115 (0.009 to 0.848)	0.012 (0.000 to 0.208)	0.000 (0.000 to 0.142)
AMI			
In-hospital mortality	0.032 (0.000 to 0.429)	0.103 (0.020 to 0.323)	0.000 (0.000 to 0.213)
30 day same diagnosis readmissions	0.000 (0.000 to 0.177)	0.000 (0.000 to 0.128)	0.499 (0.000 to 6.399)
Stroke			
In-hospital mortality	0.000 (0.000 to 0.091)	0.016 (0.000 to 0.127)	0.000 (0.000 to 0.144)

*Results expressed as SV (95% CI) are only for those indicators which showed systematic variation at the all hospital level as shown in table 4.

Table 6 Potential reductions in outcome events for diagnosis-outcome indicators*

Diagnosis-outcome indicator	Potential reductions in outcome events	
	% decrease	Events prevented No (95% CI)
Heart failure		
In-hospital mortality*	22.2%	101 (0 to 176)
Long stays	35.9%	174 (111 to 243)
All cause readmissions*	7.6%	92 (0 to 183)
Same diagnosis readmissions	22.3%	110 (48 to 175)
AMI		
In-hospital mortality	29.9%	132 (84 to 191)
Long stays*	28.9%	56 (0 to 101)
All cause readmissions*	3.7%	17 (0 to 86)
Same diagnosis readmissions	31.5%	100 (54 to 158)
Stroke		
In-hospital mortality	18.7%	115 (33 to 191)
Long stays*	9.3%	25 (0 to 84)
All cause readmissions*	15.0%	35 (0 to 77)
Same cause readmissions*	29.9%	18 (0 to 39)

*The 95% confidence intervals (CI) for events prevented include zero (that is, the reduction in events is non-significant) for indicators which did not show significant systematic variation in the all hospital analysis (table 4).

differing sample sizes and event rates, Thomas and Hofer concluded that standardised mortality estimates had very limited predictive value under even “best case” scenarios.⁴⁰

Key messages

- Reports of variation between hospitals in outcomes for specific diagnoses may be explained by the interplay between variation in data quality or reporting, differences in patient characteristics, or real variations in quality of care provided.
- For diagnosis specific outcome indicators to be useful in comparing quality of care of different hospitals and prompting local action, the level of observed variation must first be apportioned between that due to systematic variation (reflecting real variations in quality of care) and that due to case mix effects or random error in measurement using appropriate statistical methods.
- Only five of 12 diagnosis-outcome indicators relating to hospital care of patients with cardiovascular conditions were validated as profiling systematic variation between all hospitals under study, and within hospital peer groups (tertiary, community and district) only two continued to discriminate between hospitals within two hospital peer groups.
- As there is little correlation between the results of different diagnosis-outcome indicators, the overall quality of care within individual hospitals should not be inferred from the results of single or selected indicators.
- The reduction in absolute numbers of adverse outcomes that may result if all hospitals were to achieve “best performer” levels can be calculated for those indicators which, on the basis of their showing systematic variation, suggest the existence of variation in quality of care among hospitals for specific clinical conditions. Such calculations may accelerate the adoption by hospitals of condition specific quality improvement strategies.

What might be the explanations for indicator results which suggest suboptimal care? Avoidable deaths may result from the failure to provide to eligible patients those clinical interventions which randomised trials have shown to be effective in reducing mortality, or administering interventions which are harmful.⁴¹ Clearly, undertaking clinical audits using evidence based criteria of appropriateness is advisable in the setting of indicator results showing greater than expected in-hospital mortality rates.

Unexpectedly long hospital stays may result from potentially avoidable complications and/or inefficiencies in service delivery and discharge planning,⁴² particularly in the many older patients with heart failure who suffer other physical, cognitive, and psychosocial impairments. A proportion of same diagnosis readmissions is likely to be attributable to problems of incomplete resolution or stabilisation of the acute clinical syndrome during the index admission, and would again suggest the need to review in-hospital processes of care. In contrast, indicators involving all cause readmissions were not associated with systematic variation, the explanation for which may be that factors outside the control of hospitals—such as the effects of multiple comorbidities or inadequate community health care⁴³—account for many of these readmissions.

Conclusions

In conclusion, several prerequisites should be satisfied before using performance indicators based on administrative data as markers of quality of hospital care. The use of indicators should be restricted to discrete, high volume diagnoses which afford a greater likelihood of detecting systematic variation if it exists. Indicator data should be collected in a standardised manner using exclusion criteria which minimise coding and diagnostic error. Chosen indicators should be validated as markers of systematic as opposed to random variation. For those indicators which show systematic variation, quantifying and reporting the reductions in outcome events that may accrue if “best performer” standards were to be achieved universally may assist in accelerating implementation of quality improvement strategies.⁴⁴

Authors’ affiliations

I Scott, Department of Internal Medicine, Princess Alexandra Hospital, Brisbane, Queensland, Australia 4102
 D Youlden, Epidemiology Service Unit, Health Information Centre, Queensland Health Department, Brisbane, Queensland, Australia 4000

M Coory, Health Information Centre, Queensland Health Department, Brisbane, Queensland, Australia 4000

This work was performed at the Health Information Centre, Queensland Health Department, Brisbane, Queensland, Australia. 4000

REFERENCES

- 1 **Iezzoni LI**. Assessing quality using administrative data. *Ann Intern Med* 1996;**127**:666-74.
- 2 **Tu JV**, Austin P, Naylor CD, et al. Acute myocardial infarction outcomes in Ontario. In: Naylor CD, Slaughter PM, eds. *Cardiovascular health and services in Ontario: an ICES atlas*. Toronto: Institute for Clinical Evaluative Sciences, 1999:83-110.
- 3 **Jollis JG**, Romano PS. Pennsylvania's focus on heart attack: grading the scorecard. *N Engl J Med* 1998;**338**:983-7.
- 4 **Romano PS**, Zach A, Luft HS, et al. The California Hospital Outcomes Project: using administrative data to compare hospital performance. *Jt Comm J Qual Improv* 1995;**21**:668-82.
- 5 **Rosenthal GE**, Hammar PJ, Way LE, et al. Using hospital performance data in quality improvement: the Cleveland health quality choice experience. *Jt Comm J Qual Improv* 1998;**24**:347-60.
- 6 **NHS Executive**. *The new NHS, modern and dependable: a national framework for assessing performance*. London: NHS Executive, 1998.
- 7 **Australian Council of Healthcare Standards**. *Determining the potential to improve the quality of care in Australian health care organisations: results from the ACHS clinical indicators data 1998 and 1999*. Canberra: The Australian Council of Healthcare Standards, 2000.
- 8 **Gibberd RW**, Pathmeswaran A, Burtenshaw K. Using clinical indicators to identify areas for quality improvement. *J Qual Clin Practice* 2000;**20**:136-44.
- 9 **Shortell SM**, Bennett CL, Byck GR. Assessing the impact of continuous quality improvement on clinical practice: what it will take to accelerate progress. *Milbank Quarterly* 1998;**76**:593-624.
- 10 **Rainwater JA**, Romano PS, Antonius DM. The California Hospital Outcomes Project: how useful is California's report card for quality improvement? *Jt Comm J Qual Improv* 1998;**24**:31-9.
- 11 **Luce JM**, Thiel GD, Holland MR, et al. Use of risk-adjusted outcome data for quality improvement by public hospitals. *West J Med* 1996;**164**:410-4.
- 12 **Kahn KL**, Rogers WH, Rubenstein LV, et al. Measuring quality of care with explicit process criteria before and after implementation of the DRG-based prospective payment system. *JAMA* 1990;**264**:1969-75.
- 13 **Hartz AJ**, Gottlieb MS, Kuhn EM, et al. The relationship between adjusted hospital mortality and the results of peer review. *Health Serv Res* 1993;**27**:765-77.
- 14 **Iezzoni LI**, Ash AS, Shwartz M, et al. Judging hospitals by severity-adjusted mortality rates: the influence of the severity-adjustment method. *Am J Public Health* 1996;**86**:1379-87.
- 15 **Thomas JW**, Hofer TP. Research evidence on the validity of risk-adjusted mortality rate as a measure of hospital quality of care. *Med Care Res* 1998;**55**:371-404.
- 16 **Green J**, Winfield N. How accurate are hospital discharge data for evaluating effectiveness of care? *Med Care* 1993;**31**:719-31.
- 17 **Iezzoni LJ**, Foley SM, Daley J, et al. Comorbidities, complications, and coding bias. Does the number of diagnosis codes matter in predicting in-hospital mortality? *JAMA* 1992;**267**:2197-203.
- 18 **DesHarnais S**, McMahon LF, Wroblewski R. Measuring outcomes of hospital care using multiple risk-adjusted indexes. *Health Serv Res* 1991;**26**:425-45.
- 19 **Rosenthal GE**. Weak associations between hospital mortality rates for individual diagnoses: implications for profiling hospital quality. *Am J Public Health* 1997;**87**:429-33.
- 20 **Sanderson C**, McKee M. How robust are rankings? The implications of confidence intervals. *BMJ* 1998;**316**:1705.
- 21 **Wray NP**, Ashton CM, Kuykendall DH, et al. Selecting disease-outcome pairs for monitoring the quality of hospital care. *Med Care* 1995;**33**:75-89.
- 22 **Lee K**, McGreevey C. Using control charts to assess performance measurement data. *Jt Comm J Qual Improv* 2002;**28**:90-101.
- 23 **Chen J**, Radford MJ, Wang Y, et al. Performance of the '100 top hospitals': what does the report card report? *Health Aff (Millwood)* 1999;**18**:53-68.
- 24 **Chassin MR**, Park RE, Lohr KN, et al. Differences among hospitals in Medicare patient mortality. *Health Serv Res* 1989;**24**:1-8.
- 25 **Rosenthal GE**, Shah A, Way LE, et al. Variations in standardised hospital mortality rates for six common medical diagnoses. Implications for profiling hospital quality. *Med Care* 1998;**36**:955-64.
- 26 **Kiefe CI**, Weissman NW, Allison JJ, et al. Identifying achievable benchmarks of care: concepts and methodology. *Int J Qual Health Care* 1998;**10**:443-7.
- 27 **Queensland Hospitals Admitted Patients Data Collection (QHAPDC)**. *Manual of instructions and procedures for the completion of patient identification and diagnosis data*. Brisbane: Data Services Unit, Queensland Health, 2000.
- 28 **The International Classification of Diseases and Related Health Problems**, 10th revision. Australian Modification (ICD-10-AM). Sydney: National Centre for Classification in Health, University of Sydney, 2000.
- 29 **Baker GR**, Anderson GM, Brown AD, et al. *The hospital report '99: A balanced scorecard for Ontario acute care hospitals*. Toronto: Performance Management and Improvement Group, University of Toronto, 2000.
- 30 **Garnick DW**, DeLong ER, Luft HS. Measuring hospital mortality rates: are 30-day data enough? Ischaemic Heart Disease Patient Outcomes Research Team. *Health Serv Res* 1995;**29**:679-95.
- 31 **Rosenthal GE**, Harper DL, Quinn LM, et al. Severity-adjusted mortality and length of stay in teaching and nonteaching hospitals. Results of a regional study. *JAMA* 1997;**278**:485-90.
- 32 **Greenland S**. Principles of multilevel modelling. *Int J Epidemiol* 2000;**29**:158-67.
- 33 **SAS Institute Inc**. *SAS OnlineDoc*, Version 8. Cary, NC: SAS Institute, 1999.
- 34 **Hosmer DW Jr**, Lemeshow S. *Applied logistic regression*. New York: John Wiley & Sons, 1989.
- 35 **Hanley J**, McNeil B. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982;**143**:29-36.
- 36 **Coory M**, Gibberd R. New measures for reporting the magnitude of small-area variation in rates. *Stat Med* 1998;**17**:2625-34.
- 37 **Martuzzi M**, Hills M. Estimating the degree of heterogeneity between event rates using likelihood. *Am J Epidemiol* 1995;**141**:369-74.
- 38 **Goldstein H**, Spiegelhalter DJ. League tables and their limitations: statistical issues in comparisons of institutional performance. *J R Stat Soc Series A* 1996;**159**:385-443.
- 39 **Keeler EB**, Rubenstein LV, Kahn KL, et al. Hospital characteristics and quality of care. *JAMA* 1992;**268**:1709-14.
- 40 **Thomas JW**, Hofer TP. Accuracy of risk-adjusted mortality rates as a measure of hospital quality of care. *Med Care* 1999;**37**:83-92.
- 41 **Scott IA**, Harper CM. Guideline-discordant care in acute myocardial infarction: predictors and outcomes. *Med J Aust* 2002;**177**:26-31.
- 42 **Thomas JW**, Guire KE, Horvat GG. Is patient length of stay related to quality of care? *Hosp Health Serv Adm* 1997;**42**:489-507.
- 43 **Milne R**, Clarke A. Can readmission rates be used as an outcome indicator? *BMJ* 1990;**301**:1139-40.
- 44 **Kiefe CI**, Allison JJ, Williams OD, et al. Improving quality improvement using achievable benchmarks for physician feedback. A randomized controlled trial. *JAMA* 2001;**285**:2871-9.