Analysing low-risk patient populations allows better discrimination between high-performing and low-performing hospitals: a case study using in-hospital mortality from acute myocardial infarction

Michael Coory, Ian Scott

Qual Saf Health Care 2007;16:324–328. doi: 10.1136/qshc.2006.018457

Objective: To assess whether performance indicators based on administrative hospital data can be rendered more useful by stratifying them according to risk status of the patient.

Design: Retrospective analysis of 10 years of administrative hospital data for patients with acute myocardial infarction (AMI). Four risk groups defined by cross-classifying patient age (<75 years, 75+ years) against the presence or otherwise of at least one risk condition that predicted short-term AMI mortality.

Setting: 17 public hospitals in Queensland, Australia, with more than 50 AMI admissions annually.

Participants: 21 537 patients admitted through the emergency department and subsequently diagnosed as having AMI.

Main outcome measure: Systematic variation in standardised case fatality ratios. Systematic variation is the variation across hospitals after accounting for the Poisson variation in the number of deaths at each hospital. It was obtained from an empirical-Bayes model. Case fatality ratios were standardised according to the age, sex and risk factor profile of the patient.

Results: Systematic variation decreased monotonically across the four risk groups as case fatality increased (likelihood ratio test: $\chi^2 = 8.08$, $df = 1$, $p = 0.004$). Systematic variation was largest and statistically significant (0.375; 95% CI 0.144 to 0.606) for low-risk patients (<75 years with no risk conditions; case fatality rate = 2.0%) but was smallest (0.126; 0.039 to 0.212) for high-risk patients (75+ years with at least one risk condition; case fatality rate = 24.3%).

Conclusion: Analysis of data from high-risk patients with AMI provides little opportunity to identify better-performing hospitals because there is relatively little variation across hospitals. In such patients, older age and comorbid illness are probably more important than quality of care in determining outcomes. In contrast, for low-risk patients the systematic variation was large suggesting that outcomes for such patients are more sensitive to clinical error. Analysing data for low-risk patients maximises our ability to identify best-performing hospitals and learn from their processes and structures to effect system-wide changes that will benefit all patients.

See end of article for authors' affiliations

Correspondence to: Michael Coory, School of Population Health, University of Queensland, Level 3, Public Health Building, Herston Road, Herston, Queensland 4006, Australia; m.coory@sph.uq.edu.au

Accepted 2 May 2007

In spite of criticisms, there is continuing worldwide interest in using administrative hospital data to construct quality indicators.¹ The current issues are not so much about whether such indicators should be used but more about how to render them more useful.² Most of the indicators based on administrative hospital data consist of outcome-condition pairs (eg, mortality (outcome) following admission for acute myocardial infarction (condition)). Large variation across hospitals for specific indicators provides an opportunity to discern high-performing from low-performing hospitals. After identifying high-performing hospitals, their structures and processes can be analysed to formulate system-wide changes that can be applied to all hospitals to optimise quality and reduce undesirable variation. The idea is not to single out individual hospitals as "bad apples", but to identify ways of shifting all hospitals to a level of quality observed in those with the best outcomes.^{3,4} This is measurement for the purpose of learning, not judging,⁵ and embodies Demming's approach to quality improvement as applied to healthcare.⁶

The present study aimed to assess whether there were differences in the amount of variation across hospitals in quality indicators according to the risk status of the patient. If so, then stratifying by risk might be a useful approach to analysing administrative hospital data because it would

maximise the opportunity to identify best-performing hospitals. We chose the outcome-condition pair of in-hospital mortality following acute myocardial infarction (AMI) for our case study because several studies have shown a more consistent association between quality of care and AMI mortality than for some other indicators based on administrative hospital data.^{7,8}

METHODS

Patients and data

Data on patients admitted through the emergency departments of Queensland's public hospitals, who were subsequently diagnosed as having an AMI, were obtained from the Queensland Hospitals Admitted Patient Data Collection (QHAPDC). Queensland is the north-eastern state of Australia with a population of 4.0 million, which represents 18% of the total Australian population.

QHAPDC contains, inter alia, the demographic characteristics of the patients, the principal diagnosis, secondary diagnoses and the procedures performed. The data have been collected and stored in a consistent format since the financial year 1995/96 and we used the full 10 years of available data to 2004/05 to

Abbreviations: AMI, acute myocardial infarction; GLLMM, generalised latent and mixed models; SCFR, standardised case fatality ratio

obtain statistically stable estimates of across-hospital variation in mortality. Patients with AMI were identified using the *International Classification of Diseases* (ICD) version 9⁹ codes 410x (for the years 1995/96–1998/99) and ICD10¹⁰ codes I21x–I22x (1999/00–2004/05). We chose the 17 public hospitals in Queensland with more than 50 acute AMI admissions annually (range 51–253); the mean for all hospitals combined was 128.

We applied two exclusion criteria: age younger than 30 years or older than 89 years, and discharge status of alive with length of stay less than 4 days. Other studies have shown these criteria to be optimal for reducing the number of false-positive diagnoses of AMI in administrative hospital data.¹¹

Risk groups

To assess whether there was more across-hospital variation in mortality for low-risk or high-risk patients, we defined four risk groups based on patient age (<75 years, 75+ years) and the presence or otherwise of at least one risk condition identified in other studies^{11 12} (and confirmed in our data) as predicting short-term mortality for AMI (table 1).

Statistical analysis

For each of the four risk groups, we fitted logistic regression models (with 30-day in-hospital mortality as the outcome) to adjust for age, sex and risk conditions (if these were present). Risk conditions (listed in the footnote to table 1) were fitted as separate indicator variables and age was fitted to the models in 5-year groups, also using indicator variables. The c-statistics for these models ranged between 0.76 and 0.82. Statisticians generally consider that acceptable risk adjustment is indicated by a c-statistic greater than 0.70.¹³

We summed the predicted probabilities of death from the logistic regression models, within each hospital, to obtain the number of deaths (D_{exp}) that would be expected at each hospital if it had the same casemix (ie, age, sex and risk factor profile) as the average casemix for all hospitals combined.

The standardised case fatality ratio (SCFR) at each hospital was calculated as: $D_{\text{obs}}/D_{\text{exp}}$; where D_{obs} is the observed number of deaths at each hospital. SCFRs vary about 1; if a hospital has a higher mortality rate than predicted based on the age, sex and risk conditions of its patients, then its SCFR will be greater than 1. This statistical method is a standard way of constructing performance indicators that attempt to account for differences in the types of patients admitted to different hospitals.¹⁴

Estimation of the variation across hospitals should not be based directly on the observed SCFRs because these include effects of random Poisson variation within hospitals and consequently will overestimate the variation across hospitals.¹⁵ Instead, we used empirical-Bayes models to partition the variation in the observed SCFRs into: (1) that due to chance variation (ie, Poisson) in the observed number of deaths at each hospital, and (2) variation in the true underlying SCFRs, which

we have labelled as systematic variation, and was the focus of our analysis. Systematic variation has been used and validated in several studies.^{4 16}

The larger the systematic variation, the larger the true underlying variation in SCFR. In contrast, if the systematic variation is small and close to zero, then all the hospitals have approximately the same value for their true underlying SCFRs and any difference in the observed SCFRs is due to random variation in the number of deaths at each hospital.

All empirical-Bayes modelling was done in the statistical package Stata (Release 9.2) using programs which implement maximum likelihood estimation and empirical-Bayes prediction for generalised latent and mixed models (GLLAMM).¹⁷ To estimate the systematic variation for the four risk groups, we fitted four random intercept models, one for each risk group. They each had the form:

$$\ln(D_{\text{obs}}) = \ln(D_{\text{exp}}) + \beta_0 + \xi_{0j}$$

where

D_{obs} and D_{exp} are as previously defined with D_{obs} assumed to have a Poisson distribution with mean D_{obs} ;

β_0 is the mean SCFR;

$\xi_{0j} \sim N(0, SV^2)$ is the random intercept, with the j subscript corresponding to the 17 hospitals (SV , systematic variation).

The GLLAMM program provides an estimate of SV^2 and its standard error (SE). We used the delta method¹⁸ to obtain the standard error of the systematic variation ($SV = \sqrt{SV^2}$) and used this to calculate 95% confidence intervals in the usual way (ie, $1.96 \times SE$). The lower 95% confidence limit is bounded at zero because the systematic variation, like any measure of variation, cannot be negative.

To test for a statistically significant relationship between systematic variation and the risk of death, we fitted a GLLAMM to the entire dataset and specified random coefficients for risk group and a random intercept and compared it with a GLLAMM with a random intercept alone using a likelihood ratio test.¹⁹ That is, we compared the model

$$\ln(D_{\text{obs}}) = \ln(D_{\text{exp}}) + \beta_0 + \xi_{0j} + \sum (\beta_{1k} + \xi_{kj}) \cdot x_k$$

with

$$\ln(D_{\text{obs}}) = \ln(D_{\text{exp}}) + \beta_0 + \xi_{0j} + \sum \beta_{1k} \cdot x_k$$

where

x_k denotes the three indicator variables used to model the four risk groups (reference category: <75 years; no risk conditions);

Table 1 Systematic variation by risk group, 1995/96–2004/05

Risk group	Admissions n (%)	Deaths	Case fatality rate† (95% CI)‡	Systematic variation§ (95% CI)
<75 years; no risk conditions*	7452 (34.6)	148	2.0% (1.7% to 2.3%)	0.375 (0.144 to 0.606)
75+ years; no risk conditions*	2263 (10.5)	195	8.6% (7.5% to 9.8%)	0.285 (0.076 to 0.495)
<75 years; 1+ risk conditions*	6202 (28.8)	906	14.6% (13.7% to 15.5%)	0.174 (0.073 to 0.276)
75+ years; 1+ risk conditions*	5620 (26.1)	1366	24.3% (23.2% to 25.4%)	0.126 (0.039 to 0.212)
Total	21 537 (100.0)	2615	12.1% (11.7% to 12.6%)	0.184 (0.106 to 0.262)

*Dysrhythmias, conduction disorders, chronic congestive heart failure, chronic renal failure, pneumonia, chronic bronchitis and emphysema, dementia, stroke, malignancy, chronic liver failure.

†Crude case fatality per 100 admissions.

‡Exact binomial confidence interval.

§Systematic variation—that is, variation in standardised case fatality rates across hospitals after partitioning out the within hospital Poisson variation.

Figure 1 Shrunken standardised case fatality rates (SCFRs) by risk group from random intercept and random coefficient model.

β_{1k} are the mean coefficients corresponding to the indicator variables;

$\xi_{kj} \sim N(0, \psi_k^2)$ are the random coefficients, with the k subscript corresponding to the three indicator variables and the j subscript corresponding to the 17 hospitals.

The primary analysis, which was powered to definitively assess how systematic variation changed with risk status, was based on 10 years of data for all 17 hospitals combined. We also conducted secondary analyses over 2-year periods to compare the systematic variation by hospital peer group and identify best (and worst)-performing hospitals. For the analysis by hospital peer group, the four tertiary hospitals were compared with 13 community hospitals.

To identify the best-performing and worst-performing hospitals (ie, low and high outliers) we used the estimates of the underlying SCFRs from the empirical-Bayes models. Following the terminology of others,²⁰ we refer to these estimates as *shrunken SCFRs* because the effect of the empirical-Bayes models is to shrink the observed SCFRs towards the state average. The amount of shrinkage depends on the number of AMI admissions at a particular hospital. If there are fewer admissions, then there is greater shrinkage

towards the state average. The technique can be thought of as using an individual hospital's rate plus the combined rate for all hospitals to obtain a better estimate of the hospital's true underlying rate. Several authors give examples of this shrinkage, which is a way of accounting for small sample sizes at some hospitals.^{15 21–23} There is no standard, widely agreed method for using shrunken SCFRs to identify outlying hospitals.²⁴ Following Austin, we classified a hospital as a high outlier if, based on the posterior distribution, $P(\text{shrunken SCRF} > 1.15) > 0.75$; or as a low outlier if, $P(\text{shrunken SCRF} < 0.85) > 0.75$.²⁴ Others have also used this method.¹⁵

RESULTS

There were 21 537 AMI admissions to the 17 largest public hospitals in Queensland during the 10-year study period that met the selection criteria. During this time the annual number of AMI admissions remained relatively stable at 2071 in 1995/96 and 2234 in 2004/05, while the crude case fatality rate decreased from 13.7% to 10.4%; the combined rate for all years was 12.1%.

The case fatality rate varied greatly across the four risk groups: from 2.0% for AMI patients younger than 75 years with no risk conditions to 24.3% for AMI patients older than 75 years with a least one risk condition (table 1). The point estimate of the systematic variation was largest (0.375) for the group with the lowest (2.0%) case fatality rate (ie, <75 years and no risk conditions). The systematic variation in the three other groups monotonically decreased as the case fatality rate increased (fig 1). This inverse relationship between systematic variation and the risk of death was statistically significant ($\chi^2 = 8.08$, $df = 1$, $p = 0.004$, based on likelihood ratio test comparing the model with random coefficients for risk group and a random intercept with the model with a random intercept alone).

For the secondary analysis of hospital peer groups for the two most recent years, systematic variations in each of the four risk groups were so imprecise (extremely wide confidence intervals) as to not provide any useful information. Therefore, we collapsed the four risk groups into two according to whether the patient had at least one risk condition. For all hospitals combined (last three rows in table 2), systematic variation for the lower-risk group was 0.359 (95% CI 0.087 to 0.632), and was larger than that for the higher-risk group—0.121 (95% CI 0.000 to 0.285), although the difference between groups was not statistically significant ($p = 0.168$). Similarly, when stratified by peer group the data showed the same pattern of higher

Table 2 Systematic variation by collapsed risk groups and hospital peer groups, 2003/04–2004/05

Risk group/hospital peer group	Admissions n (%)	Deaths	Case fatality† (95% CI)‡	Systematic variation§ (95% CI)
Tertiary				
No risk conditions	671 (46.4)	15	2.2% (1.3% to 3.7%)	0.468 (0.000 to 1.161)
At least one risk condition*	775 (53.6)	137	17.7% (15.1% to 20.5%)	0.000 (0.000 to 0.196)
All patients	1446 (100.0)	152	10.5% (9.0% to 12.2%)	0.109 (0.000 to 0.353)
Base				
No risk conditions	1549 (48.2)	50	3.2% (2.4% to 4.2%)	0.241 (0.000 to 0.797)
At least one risk condition*	1663 (51.8)	287	17.3% (15.5% to 19.2%)	0.128 (0.000 to 0.329)
All patients	3212 (100.0)	337	10.5% (9.5% to 11.6%)	0.187 (0.036 to 0.338)
All hospitals				
No risk conditions	2220 (47.7)	65	2.9% (2.3% to 3.7%)	0.359 (0.087 to 0.632)
At least one risk condition*	2438 (52.3)	424	17.4% (15.9% to 19.0%)	0.121 (0.000 to 0.285)
All patients	4658 (100.0)	489	10.5% (9.6% to 11.4%)	0.203 (0.064 to 0.342)

*Dysrhythmias, conduction disorders, chronic congestive heart failure, chronic renal failure, pneumonia, chronic bronchitis and emphysema, dementia, stroke, malignancy, chronic liver failure.

†Crude case fatality per 100 admissions.

‡Exact binomial confidence interval.

§Systematic variation—that is, variation in standardised case fatality rates across hospitals after partitioning out the within hospital Poisson variation.

Table 3 Low and high outliers for collapsed low-risk group* and all patients combined, 1995/96–2004/05

Period	Low-risk patients			All patients combined		
	Outliers			Outliers		
	Low	Neither	High	Low	Neither	High
1995/96–1996/97	1	15	1	0	17	0
1997/98–1998/99	2	12	3	1	13	3
1999/00–2000/01	1	13	3	0	15	2
2001/02–2002/03	1	15	1	1	14	2
2003/04–2004/05	2	13	2	1	15	1

See text for concordance between outliers for low-risk patients and all patients combined.

*Patients with no reported risk conditions.

systematic variation in the lower risk group, although with smaller sample sizes we could not show these results to be statistically significant (table 2).

For the analysis of outliers we also collapsed the two low-risk groups into one (table 3). More outliers were identified using the collapsed low-risk group than using all patients combined. Of the 17 outliers identified in the low-risk group, 7 (41%) were also identified in the combined group; these 7 concordant outliers represented 64% of the 11 outliers identified for all patients combined.

DISCUSSION

There was more variation in mortality across hospitals for low-risk than for high-risk patients with AMI. Analysis of data for low-risk patients might therefore facilitate the process of identifying best-performing hospitals and learning from their structures and processes to effect system-wide change that will benefit all patients.

More low or high outliers were identified using the low-risk group than using all patients combined. This was not surprising because the larger systematic variation for the low-risk group means that even after taking into account the random variation related to the number of deaths at each hospital, there was still more variation across hospitals and consequently more outliers in the low-risk group than in the other groups.

What is already known about this topic

- There is continuing interest in assessing hospital performance using diagnosis-specific quality indicators based on administrative data.
- Screening large amounts of administrative data is an inexpensive method for identifying variations across hospitals which might warrant more detailed clinical analysis.

What this study adds

- There is less variation across hospitals in mortality for high-risk than low-risk patients admitted for acute myocardial infarction.
- Analysis of data for low-risk patients maximises our ability to identify best-performing hospitals and learn from their processes and structures to effect system-wide changes that will benefit all patients.

We could find only one other published analysis of variation across hospitals by risk status: a Canadian database study of AMI mortality.²⁵ The main focus of that study was to compare the performance of bayesian and frequentist methods in identifying outliers. In a secondary analysis, the authors used bayesian hierarchical models to specify three fictitious patient profiles: (1) low risk: a man aged 50–64 years, with no additional risk factors, (2) medium risk: a patient aged 65–74 years, all of whose risk factors were set to the cohort average, (3) high risk: a woman aged at least 75 years with chronic renal failure, heart failure and cardiac dysrhythmias. For each risk profile, the probability of mortality at each of 139 hospitals was calculated. Similar to our results, they found the greatest across-hospital variation for the low-risk group.

Study limitations

Our analyses did not address any of the possible systematic errors (biases) that can occur in analyses of large administrative databases.²⁶ For example, we had information on age, sex, heart failure and dysrhythmias, but we did not have information on potentially important variables such as smoking or obesity. Also, low-risk patients were defined as those with no reported risk conditions and it is possible that some of these patients had unreported risk conditions. Also, there might be differences across hospitals in the diagnostic threshold for AMI, especially with the more widespread use of troponin assays.²⁷ Because of problems such as these, analyses of administrative data are best viewed as a way of screening large routinely maintained data systems; they are not definitive.²⁶ Instead, their aim is to suggest avenues for more indepth analyses and use the lessons subsequently learnt to improve quality of care. Along these lines, Mohammed and co-workers have suggested the pyramid of investigation model, which considers data problems and residual differences in casemix (as discussed in the preceding paragraph) before attributing any variation across hospitals to variation in quality of care.²⁸ Despite these limitations, hospital-specific quality indicators are often based on administrative data² and we were interested in how to maximise the usefulness of such data.

Another limitation of the present study is that a large sample size (ie, 10 years of data) was required to properly assess whether systematic variation decreased monotonically as risk increased across the four groups. The aim of routine quality measurement is to instigate timely remedial action in clinical care and a shorter study period is preferred. Our secondary analyses, based on 2 years of data, showed relatively more variation for the low-risk group.

Implications for quality improvement programmes

Medical errors are more prevalent among elderly patients, who are frail and more likely to have complicating, comorbid conditions.²⁹ Although a potentially attractive strategy for reducing clinical error is to target patients at high risk of adverse outcomes, it does not necessarily follow that we have the most to learn about improving systems of care by analysing data for this group exclusively. There was so little variation across hospitals for high-risk patients that analysing this patient subpopulation would afford little opportunity for learning to improve quality. It is likely that for high-risk patients, factors such as advanced age and comorbid conditions overwhelm variation in quality of care in determining outcomes. In contrast, low-risk patients showed the greatest variation in mortality across hospitals, suggesting that the outcome for this patient group is particularly sensitive to the prevalence of clinical error. Analysing quality indicators for low-risk patients might therefore afford a better chance of

identifying and learning from hospitals which have better outcomes.

This idea is an extension of the idea behind quality indicators based on diagnosis-related groups (DRGs) with extremely low mortality.³⁰ Applying this idea to a particular condition (ie, AMI), rather than a group of heterogeneous DRGs, might provide more specific opportunities for learning.

CONCLUSION

Analysis of administrative hospital data after first stratifying by patient risk status is a more useful method for identifying best-performing hospitals than using quality indicators based on all patients combined. Learning from best-performing hospitals identified in this manner might contribute substantially to our knowledge and ability to improve care for all patients.

AUTHORS' CONTRIBUTIONS

MC did the statistical analysis. IS conceived the study. Both authors wrote, edited and critically reviewed the manuscript. MC is the guarantor.

Authors' affiliations

Michael Coory, School of Population Health, University of Queensland, Queensland, Australia
Ian Scott, Clinical Services Evaluation Unit, Princess Alexandra Hospital, University of Queensland, Queensland, Australia

Funding: This work was completed using resources routinely available at the authors' workplace. There was no external funding. The authors' employers had no role in the study design, analysis or interpretation.

Competing interest: None.

Ethical approval: The study was based on administrative hospital data and there was no contact with patients. No information that could identify an individual patient was used in the analysis. The Queensland Department of Health, the data custodian, did not require ethics approval for the study.

REFERENCES

- 1 **Powell A**, Davies H, Thomson R. Using routine comparative data to assess the quality of health care: understanding and avoiding common pitfalls. *Qual Saf Health Care* 2003;**12**:122–8.
- 2 **Marshall M**, Romano P. Impact of reporting hospital performance. *Qual Saf Health Care* 2005;**14**:77–8.
- 3 **Brophy J**. Practice variations, chance and quality of care. *CMAJ* 1998;**159**:949–52.
- 4 **Howley P**, Gibberd R. Using hierarchical models to analyse clinical indicators: a comparison of the gamma-Poisson and beta-binomial models. *Int J Qual Health Care* 2003;**15**:319–29.
- 5 **Berwick D**. A primer on leading the improvement of systems. *BMJ* 1996;**312**:619–22.
- 6 **Demings W**. *The new economics*. Boston, MA: MIT Press, 1993.
- 7 **Scott I**, Coory M, Harper C. The effects of quality improvement interventions on in-hospital mortality after acute myocardial infarction. *Med J Aust* 2001;**175**:465–70.
- 8 **Jollis J**, DeLong E, Peterson E, *et al*. Outcome of acute myocardial infarction according to speciality of admitting physician. *N Engl J Med* 1996;**335**:1880–7.
- 9 **National Coding Centre**. Australian version of the International Classification of Diseases, 9th Revision, Clinical Modification (ICD-9-CM). Sydney: National Coding Centre, University of Sydney, 1996.
- 10 **National Centre for Classification in Health**. The International Statistical Classification of Diseases and Related Health Problems, 10th Revision, Australian Modification. Sydney: NCCH, University of Sydney, 2002.
- 11 **Tu J**, Austin P, Naylor C, *et al*. Acute myocardial infarction outcomes in Ontario. In: Naylor C, Slaughter P, eds. *Cardiovascular health and services in Ontario: an ICES Atlas*. Toronto: Institute for Clinical Evaluative Sciences, 1999:83–100.
- 12 **Alter D**, Naylor C, Austin P, *et al*. Effects of socioeconomic status on access to invasive cardiac procedures and on mortality after acute myocardial infarction. *N Engl J Med* 1999;**341**:1359–67.
- 13 **Hosmer D**, Lemeshow S. *Applied logistic regression*. New York: John Wiley & Sons, 1989.
- 14 **Hosmer D**, Lemeshow S. Confidence interval estimates of an index of quality performance based on logistic regression models. *Stat Med* 1995;**14**:2161–72.
- 15 **Christiansen C**, Morris C. Improving the statistical approach to health care provider profiling. *Ann Intern Med* 1997;**127**:764–8.
- 16 **Coory M**, Gibberd R. New measures for reporting the magnitude of small area variation in rates. *Stat Med* 1998;**17**:2625–34.
- 17 **Rabe-Hesketh S**, Skrondal A, Pickles A. *GLLAMM manual*, U.C.Berkeley Division of Biostatistics working paper series. Berkeley: Berkeley Electronic Press, 2004.
- 18 **Kirkwood B**, Sterne J. *Essential medical statistics*. Oxford: Blackwell Science, 2005.
- 19 **Rabe-Hesketh S**, Pickles A, Skrondal A. Reliable estimation of generalized linear mixed models using adaptive quadrature. *Stata Journal* 2002;**2**:1–21.
- 20 **Gibberd R**, Hancock S, Howley P, *et al*. Using indicators to quantify the potential to improve quality of care. *Int J Qual Health Care* 2004;**16**:i37–43.
- 21 **Clayton D**, Kaldor J. Empirical Bayes estimates of age-standardized relative risk for use in disease mapping. *Biometrics* 1987;**43**:671–81.
- 22 **Spiegelhalter D**, Myles J, Jones D, *et al*. An introduction to bayesian methods in health technology assessment. *BMJ* 1999;**319**:508–12.
- 23 **Greenland S**. Principles of multilevel modelling. *Int J Epidemiol* 2000;**29**:158–67.
- 24 **Austin P**. A comparison of Bayesian methods for profiling hospital performance. *Med Dec Making* 2002;**22**:163–72.
- 25 **Austin P**, Naylor C, Tu J. A comparison of a Bayesian versus a frequentist method for profiling hospital performance. *J Eval Clin Prac* 2001;**7**:35–45.
- 26 **Iezzoni L**. Assessing quality using administrative data. *Ann Intern Med* 1997;**127**:666–74.
- 27 **Kavsak P**, MacRae A, Lustig V, *et al*. The impact of the ESC/ACC redefinition of myocardial infarction and new sensitive troponin assays on the frequency of acute myocardial infarction. *Am Heart J* 2006;**152**:118–25.
- 28 **Mohammed A**, Rathbone A, Myers P, *et al*. An investigation into general practitioners associated with high patient mortality flagged up through the Shipman inquiry: retrospective analysis of routine data. *BMJ* 2004;**328**:1474–7.
- 29 **Weingart S**, Wilson R, Gibberd R, *et al*. Epidemiology of medical error. *BMJ* 2000;**320**:774–7.
- 30 **Agency for Healthcare Quality and Research**. AHRQ quality indicators. Guide to patient safety indicators [version 2.1, revision 3]. Rockville, MD: AHRQ, 2005.