Does the choice of risk-adjustment model influence the outcome of surgeon-specific mortality analysis? A retrospective analysis of 14 637 patients under 31 surgeons

S W Grant, A D Grayson, M Jackson, J Au, B M Fabri, G Grotte, M Jones, B Bridgewater

University Hospital of South Manchester NHS Foundation Trust, Southmoor Road, Manchester M23 9LT, UK

Correspondence to: Mr Ben Bridgewater, University Hospital of South Manchester NHS Foundation Trust, Southmoor Road, Manchester M23 9LT, UK; ben.bridgewater@smuht.nwest.nhs.uk

On behalf of the North West Quality Improvement Programme in Cardiac Interventions (NWQIP)

Accepted 9 October 2007

ABSTRACT

Objectives: To compare implications of using the logistic EuroSCORE and a locally derived model when analysing individual surgeon mortality outcomes.

Design: Retrospective analysis of prospectively collected data.

Setting: All NHS hospitals undertaking adult cardiac surgery in northwest England.

Patients: 14 637 consecutive patients, April 2002 to March 2005.

Main outcome measures: We have compared the predictive ability of the logistic EuroSCORE (uncalibrated), the logistic EuroSCORE calibrated for contemporary performance and a locally derived logistic regression model. We have used each to create risk-adjusted individual surgeon mortality funnel plots to demonstrate high mortality outcomes.

Results: There were 458 (3.1%) deaths. The expected mortality and receiver operating characteristic (ROC) curve values were: uncalibrated EuroSCORE —5.8% and 0.80, calibrated EuroSCORE —3.1% and 0.80, locally derived model —3.1% and 0.82. The uncalibrated EuroSCORE plot showed one surgeon to have mortality above the northwest average, and no surgeon above the 95% control limit (CL). The calibrated EuroSCORE plot and the local model showed little change in surgeon ranking, but significant differences in identifying high mortality outcomes. Two of three surgeons above the 95% CL using the calibrated EuroSCORE revert to acceptable outcomes when the local model is applied but the finding is critically dependent on the calibration coefficient.

Conclusions: The uncalibrated EuroSCORE significantly overpredicted mortality and is not recommended. Instead, the EuroSCORE should be calibrated for contemporary performance. The differences demonstrated in defining high mortality outcomes when using a model built for purpose suggests that the choice of risk model is important when analysing surgeon mortality outcomes.

Results of cardiac surgical operations are now subject to close scrutiny in the United Kingdom. Mortality rates of individual surgeons have been published by the media,¹ through peer review publication,² on a Healthcare Commission website and through hospital internet sites.³ Supporters of this initiative believe that publication will help to improve quality, enable patient choice and provide reassurance that no hospital or surgeon has excessive mortality.⁴

Various surgical factors are important determinants of operative mortality including the type of operation and various patient risk factors.⁵ Unless these factors are taken into consideration it is possible to obtain false conclusions from the comparison of mortality rates, and any initiative that examines surgical outcomes without adjusting for predicted risk may lead the surgical community to turn down high-risk patients who might otherwise benefit from surgery.^{6–10} It has been clearly shown that the proportion of high-risk cases performed by individual surgeons can differ markedly, meaning that potential pitfalls of using non-risk-adjusted data are real.¹¹

The most commonly used tool to adjust for operative risk is the EuroSCORE, which has two variants; a simple additive EuroSCORE and a more complex logistic version.^{12–13} The additive EuroSCORE has been shown to overpredict operative risk and shows poor predictive ability in higher-risk patients.^{11–14–15} The logistic EuroSCORE has also been shown to overpredict observed mortality but has better predictive ability for high-risk patients.¹⁶ The logistic EuroSCORE has been used for risk adjustment in recent analyses on surgical outcomes by the Healthcare Commission in the UK.³ This type of analysis will become increasingly important when the recommendations of the Department of Health white paper “Trust, assurance and safety”, of using clinical outcomes data for professional recertification are implemented.¹⁷

It is not clear what influence the selection of different risk-adjustment models has on the analysis of individual surgical outcomes. To study this in more detail we have used a large regional patient database to derive the “best” statistical model possible based on local data. We have then compared surgeon-specific mortality analyses using crude mortality, the logistic EuroSCORE in its original form, the logistic EuroSCORE calibrated for contemporary performance and our locally derived purpose-built model.

METHODS

The North West Quality Improvement Programme in Cardiac Interventions (NWQIP) is a regional consortium involving all four NHS hospitals performing adult cardiac surgery and percutaneous coronary interventions in the north west of England (Blackpool Victoria Hospital,

Table 1 Cardiac operations performed

	Number of patients	% of activity	In-hospital mortality (%)
Isolated CABG	9671	66.1	1.9
Isolated valve	2311	15.8	3.3
CABG + valve	1502	10.3	6.7
CABG + other	207	1.4	5.8
Valve + other	325	2.2	6.5
CABG + valve + other	119	0.8	11.8
Other	502	3.4	10.0

CABG, coronary artery bypass grafting.

Other, any other major cardiac procedure excluding isolated CABG, isolated valve procedure and combined CABG and valve procedure.

Cardiothoracic Centre-Liverpool, Manchester Royal Infirmary and South Manchester University Hospital).¹⁸

Data were collected prospectively on 14 637 consecutive patients undergoing adult cardiac surgery between 1 April 2002 and 31 March 2005. Each patient had a dataset collected, which included preoperative and operative variables, to enable a predicted mortality to be calculated. Data were collected in each institution and returned to a central source for analysis. Validation of data was conducted in each centre. Mortality was defined as any in-hospital death.

Design of the study

The specific questions we addressed were:

- Are there differences in the conclusions drawn when using the various models for analysing individual surgeon data?
- Is there a difference in predictive ability between the models?
- Are the significant patient risk factors found in the local purpose-built risk-prediction model different from the risk factors included in the logistic EuroSCORE?

Table 2 Independent risk factors for in-hospital mortality ($p < 0.001$ for all variables)

	Odds ratio (95% CI)
Age (years)	1.04 (1.03 to 1.05)
Female sex	1.8 (1.5 to 2.3)
Respiratory disease	1.6 (1.3 to 2.0)
Peripheral vascular disease	1.7 (1.3 to 2.2)
Renal dysfunction	3.9 (2.9 to 5.2)
Endocarditis	5.1 (2.5 to 10.6)
Ejection fraction <30%	2.4 (1.9 to 3.2)
Previous cardiac surgery	2.6 (1.9 to 3.6)
Urgent surgery	2.4 (2.0 to 3.0)
Emergency surgery	4.3 (2.9 to 6.5)
Cardiogenic shock	4.9 (3.1 to 7.8)
Combined CABG and valve surgery	2.3 (1.8 to 2.9)
Other cardiac surgery	2.6 (2.0 to 3.4)

CABG, coronary artery bypass grafting.

Calculation of predicted risk using patient data and logistic regression coefficients:

Calculate the odds of in-hospital death = $\exp(-7.4554 + [0.0393 \times \text{age (continuous variable)}] + [0.6108 \times \text{female sex}] + [0.4996 \times \text{respiratory disease}] + [0.5305 \times \text{peripheral vascular disease}] + [1.3592 \times \text{renal dysfunction}] + [1.6313 \times \text{endocarditis}] + [0.8915 \times \text{ejection fraction <30\%}] + [0.9731 \times \text{previous cardiac surgery}] + [0.8878 \times \text{urgent surgery}] + [1.4697 \times \text{emergency surgery}] + [1.5900 \times \text{cardiogenic shock}] + [0.8282 \times \text{combined CABG and valve surgery}] + [0.9629 \times \text{other cardiac surgery}])$.
Predicted risk of in-hospital mortality as a percentage = $[\text{odds}/(1 + \text{odds})] \times 100$.

Statistical analysis

Categorical data are shown as a percentage. Predicted mortality was calculated for each patient by using the logistic EuroSCORE formula.¹⁹ If a patient factor necessary to calculate the EuroSCORE was missing in the record, that factor was assumed to be absent (occurred in less than 2% of cases).

A multivariate logistic regression analysis was undertaken on the data, using the forward stepwise technique, to identify independent risk factors for in-hospital mortality.²⁰ Candidate variables were entered into the model with a p value less than 0.1. The predicted risks of individual patients were rank ordered and divided into 10 groups. Within each group of estimated risk, the number of in-hospital deaths predicted was compared with the number of observed in-hospital deaths and the Hosmer-Lemeshow goodness-of-fit statistic was calculated to assess the calibration of the model.²⁰ The area under the receiver operating characteristic (ROC) curve was calculated to assess discriminatory ability of both the local model and the EuroSCORE.²¹

Observed mortality was compared to expected mortality from the logistic EuroSCORE (uncalibrated), a regionally calibrated logistic EuroSCORE and the locally derived logistic model. The regionally calibrated logistic EuroSCORE was derived by dividing the observed percentage mortality by the percentage predicted by the uncalibrated logistic EuroSCORE and this was done in two ways: first, by using the figures published in our most recent analysis of data during the period 2002–4, which produces a calibrated logistic EuroSCORE set at 0.58 of the original.¹⁶ Second, we have simply divided the observed mortality by that predicted by logistic EuroSCORE in the current study—that is, $3.1\% \text{ divided by } 5.8\% = 0.53$.

To assess the difference in risk-factor weightings between models we calculated the odds ratio from the logistic EuroSCORE coefficients and compared these to the odds ratio values of the locally derived model.

To assess the effect of using different risk models in identifying high mortality and potential outliers we have used funnel plots with 95% and 99% control limits.²² Risk-adjusted mortality was calculated using the uncalibrated logistic EuroSCORE, the calibrated logistic EuroSCORE and the locally derived model. This method of risk adjustment was developed by Hannan *et al.*,²³ and is derived by calculating the observed and expected mortality ratios for each surgeon. These ratios are multiplied by the region-wide observed mortality. The result reflects performance expected had the case-mix of the surgeon been identical to that of the region.

To examine how different risk-adjustment methods affect rank ordering of surgeons we divided surgeons into quartiles based on observed mortality. We then ranked surgeons using the different risk-adjustment methods and examined the effect each method had on surgeon rank and quartile. For a closer examination of how risk adjustment using a local model would differ from using the logistic EuroSCORE, the surgeon rankings based on the local model and logistic EuroSCORE were correlated using Spearman's rank correlation.

All analysis was performed using SAS for Windows Version 8.2.

RESULTS

In-hospital mortality rate and breakdown of operations

There were 458 in-hospital deaths in 14 637 patients giving a mortality of 3.1%. The operation type and in-hospital mortality rates are listed in table 1.

Table 3 Expected in-hospital mortality rates

Hospital	Observed in-hospital mortality (%)	Uncalibrated Logistic EuroSCORE (%)	Calibrated logistic EuroSCORE 2002–4 (%)	Calibrated logistic EuroSCORE 2002–5 (%)	NWQIP model (%)
All	3.1	5.8	3.4	3.1	3.1
1	3.1	5.5	3.2	2.9	3.3
2	3.6	6.8	3.9	3.6	3.4
3	3.1	5.1	3.0	2.7	3.0
4	2.6	5.2	3.1	2.8	3.0

EuroSCORE, European System for Cardiac Operative Risk Evaluation; NWQIP, North West Quality Improvement Programme.

Independent risk factors for in-hospital mortality

The independent risk factors for mortality, along with odds ratios and confidence limits are shown in table 2. The area under the ROC curve for the locally derived multivariate prediction model was 0.82. The Hosmer-Lemeshow goodness-of-fit statistic across groups of risk was not statistically significant ($p = 0.33$), indicating no evidence for a lack of fit. The logistic regression equation for calculation of predicted risk of in-hospital mortality is shown at the bottom of table 2.

Performance of different risk models

The expected hospital mortality rates from the different risk models are shown in table 3. The area under the ROC curve for both the uncalibrated logistic EuroSCORE and the calibrated logistic EuroSCORE was 0.80.

Difference in risk factors between EuroSCORE and the NWQIP model

Table 4 highlights the differences in risk factors between the logistic EuroSCORE and the NWQIP model and assesses whether the weightings given are different. The EuroSCORE calculations were based on the original risk factor definitions; however, the definitions for a number of risk factors included in the NWQIP model are different. Renal dysfunction in the NWQIP model includes both patients with a serum creatinine

>200 $\mu\text{mol/l}$ and patients who are currently receiving dialysis. In the NWQIP model, the risk factor “critical preoperative state” is only present if the patient is in cardiogenic shock. In the EuroSCORE the risk factor “other than isolated CABG” includes any other major cardiac procedure other than, or in addition to, CABG. In the NWQIP model there are two risk factors involving surgery other than isolated CABG. The first risk factor in the model is CABG combined with a valve procedure and the second risk factor is having any other surgery not including CABG, valve repair/replacement or a combined CABG and valve procedure. This risk factor in the local model therefore includes surgery on the thoracic aorta and surgery for post-infarct septal rupture, which are included as risk factors in the EuroSCORE. Any risk factors noted as not in model (NIM) in table 4 were offered to the logistic regression model but were not identified as independent risk factors.

Funnel plots

Figure 1 shows observed and risk-adjusted mortality funnel plots with 95% and 99% control limits. The observed in-hospital mortality funnel plot (fig 1A) showed one surgeon outside the upper 99% control limit, and two additional surgeons outside the upper 95% control limit. Two of these three surgeons return to satisfactory outcomes when the local model is used (inside the 95% control limits), and the other remains between the 99%

Table 4 Differences in risk factors

Risk factor	EuroSCORE coefficient	EuroSCORE odds ratio	NWQIP model odds ratio (95% CI)
Age	0.0666354	1.07	1.04 (1.03 to 1.05)
Female sex	0.3304052	1.4	1.8 (1.5 to 2.3)
Chronic pulmonary disease	0.4931341	1.6	1.6 (1.3 to 2.0)
Extracardiac arteriopathy	0.6558917	1.9	1.7 (1.3 to 2.2)
Neurological dysfunction disease	0.841626	2.3	NIM
Previous cardiac surgery	1.002625	2.7	2.6 (1.9 to 3.6)
Serum creatinine	0.6521653	1.9	3.9 (2.9 to 5.2)
Active endocarditis	1.101265	3.0	5.1 (2.5 to 10.6)
Critical preoperative state	0.9058132	2.5	4.9 (3.1 to 7.8)
Unstable angina	0.5677075	1.8	NIM
LV dysfunction moderate	0.4191643	1.5	NIM
LV dysfunction severe	1.094443	3.0	2.4 (1.9 to 3.2)
Recent myocardial infarct	0.5460218	1.7	NIM
Pulmonary hypertension	0.7676924	2.2	NIM
Urgent surgery	NIM	NIM	2.4 (2.0 to 3.0)
Emergency surgery	0.7127953	2.0	4.3 (2.9 to 6.5)
Other than isolated CABG	0.5420364	1.7	2.3 (1.8 to 2.9) *
			2.6 (2.0 to 3.4) †
Surgery on thoracic aorta	1.159787	3.2	2.6 (2.0 to 3.4) †
Postinfarct septal rupture	1.462009	4.3	2.6 (2.0 to 3.4) †

EuroSCORE, European System for Cardiac Operative Risk Evaluation; NWQIP, North West Quality Improvement Programme; CI, confidence interval; NIM, not in model; LV, left ventricular; CABG, coronary artery bypass grafting.

*Combined CABG and valve surgery in NWQIP model.

†Other cardiac surgery in NWQIP model.

Figure 1 Funnel plots for surgeon-specific observed in-hospital mortality (A), risk adjusted in-hospital mortality using uncalibrated logistic EuroSCORE (B), risk adjusted in-hospital mortality using NWQIP model (C), risk adjusted in-hospital mortality using calibrated logistic EuroSCORE (2002–5) (D), and risk adjusted in-hospital mortality using calibrated logistic EuroSCORE (2002–4) (E).

and 95% control limits. The risk-adjusted mortality funnel plot using uncalibrated logistic EuroSCORE (fig 1B) shows all surgeons except one to have mortality below the regional average and no surgeon above the 95% control limit.

The risk-adjusted mortality funnel plot for the local model (fig 1C) and the calibrated logistic EuroSCORE 2002 to 2005 (fig 1D) both show no surgeon to be above the 99% control limit. The calibrated EuroSCORE 2002–5 plot shows three surgeons falling between the 95% and 99% control limit. When the local model is used two of these three surgeons return to satisfactory outcomes. The local model continues to identify a middle-volume surgeon at between 95% and 99% control limits and, in addition, also identifies a low-volume surgeon to be between 95% and 99% control limits. When the funnel plots of the calibrated EuroSCORE 2002 to 2005 (fig 1D) and the calibrated EuroSCORE 2002–4 (fig 1E) are compared it can be

seen that two of the three outliers at 95% control limits return to “acceptable performance” because of the more lenient calibration coefficient used.

Quartiles

Rank quartile analysis of crude mortality data showed eight surgeons in the highest mortality quartile. Following risk adjustment two of these surgeons drop out of the highest mortality quartile and are moved to the second highest quartile with two surgeons moving in the other direction. The logistic EuroSCORE and the locally derived model defined the same eight surgeons in both the highest mortality and the lowest mortality quartiles; however there was some movement of surgeons within the quartiles. Figure 2 shows that the local model and the logistic EuroSCORE were highly correlated with respect to surgeon ranking.

Figure 2 Local model surgeon ranking versus logistic EuroSCORE surgeon ranking.

DISCUSSION

Statement of principal findings

This study has shown that it is possible to derive a local risk-adjustment model built for purpose that has better predictive ability than the logistic EuroSCORE. There are a number of risk factors in the local model that are not included in the EuroSCORE and vice versa. In addition, a number of the significant risk factors have different weightings in the locally derived model than in the EuroSCORE. Analysis of surgical outcomes using funnel plot methodology has shown that there are outliers on crude mortality plots that revert to acceptable outcomes following appropriate risk adjustment. Use of an uncalibrated logistic EuroSCORE plot gives false reassurance about outcomes and should not be used. On the whole the conclusions reached when using either the calibrated logistic EuroSCORE or the purpose-built locally derived model are similar and the correlation between individual surgeon ranking is strong. However, which surgeons are defined as showing “outlying” performance is critically dependent on the exact time period used to calibrate the EuroSCORE, as there have been ongoing improvements in overall mortality with time.

Strengths and weaknesses of the study

Our study has been conducted on a large patient population undergoing surgery over three years in northwest England and includes all patients undergoing adult cardiac surgery in NHS hospitals equating to about one eighth of all cardiac surgical activity in the UK. We have shown previously that outcomes in the north west are similar to those seen nationally.¹⁶ The data have the confidence of clinicians and are locally validated at each institution but have not been externally validated, which is a weakness of the study. The size of the study sample is large and so we think our findings are robust.

Strength and weaknesses compared to other studies

Risk stratification is important when comparing outcomes in cardiac surgery. It has been shown that the predicted mortality differs significantly between surgeons owing to the number of high-risk cases taken on by each surgeon.¹¹ If published outcomes are compared using only crude mortality this may lead to high-risk cases being turned down to protect results.^{6–10} This study shows clearly that non-adjusted analyses may be misleading: the one surgeon falling outside 99% control limits

on the “crude” mortality plot falls within the confidence intervals following adjustment with either the calibrated (2002–5) or locally derived model.

The first risk prediction tool to be widely used in cardiac surgery was developed in the 1980s.²⁴ It was used to compare risk-adjusted outcomes but was subsequently shown to significantly overpredict observed mortality.^{5, 25} The EuroSCORE was created to produce an objective model based on a wide patient population. It was developed using a European database of 19 030 consecutive patients undergoing cardiac surgery at 128 surgical centres in eight European countries throughout the 1990s.¹² Data were collected on 68 preoperative and 29 objective operative risk factors and then analysed comparing the risk factors to patient mortality outcome using logistic regression.

The EuroSCORE was first used in 1999 as a simple additive model and has since been extensively studied.^{12–14} The studies have shown that the additive EuroSCORE overestimates mortality in low-risk patients and underestimates mortality in high-risk patients as well as underestimating mortality in combined CABG and valve procedures.^{11–14, 15} To try to improve the performance of the additive EuroSCORE, the logistic EuroSCORE model was recommended.^{13, 26} However recent studies have shown the logistic EuroSCORE to overpredict observed mortality in all operative groups and in both high-risk and low-risk patients.^{16–27}

We are not aware of previous studies analysing the implication of using different models on defining outlying performance

Meaning of the study

The models

This study defines in more detail some of the limitations of the logistic EuroSCORE. A number of risk factors such as unstable angina and recent myocardial infarction, which are included as significant factors in the EuroSCORE, are not shown to be significant risk factors in our model, which probably reflects different models of care with increased use of routine interventional strategies for acute coronary syndromes and better overall outcomes for these patients in the modern era. Moderate left ventricular dysfunction is also included in the EuroSCORE but not in our model and this again probably indicates improved care for this group of patients either with better myocardial protection or off-pump surgery. Age as a risk factor has a lower weighting in our model than the EuroSCORE; for example, an otherwise fit 83-year-old patient for isolated first-time coronary surgery would have a predicted mortality using the original logistic EuroSCORE of 4.2%, but using the locally derived model the predicted mortality is now 1.8%. This is probably the major reason why the EuroSCORE has “drifted” and now requires recalibration.

This study has shown that if you risk adjust using the calibrated EuroSCORE or the locally derived model then different surgeons may be identified as outliers at 95% control limits. Being defined as an outlier on a funnel plot does not mean that a surgeon is “performing badly”. Issues around multiple comparisons mean that you would expect at least one surgeon out of 31 to have outlying performance at 95% control limits by chance alone. However being identified as an outlier may have a significant impact on an individual surgeon and should lead to a deeper investigation of practice involving case-mix review and a more comprehensive analysis of the workplace and associated issues. In this case, using 95% control limits, two surgeons would avoid further analysis and one surgeon would require further analysis if a model built for purpose was used instead of the calibrated logistic EuroSCORE.

Previous published work from our group clearly showed that the logistic EuroSCORE overpredicts observed mortality on a dataset collected in 2002 to 2004 and suggested a calibration factor of 0.58 was necessary to adjust the logistic EuroSCORE for contemporary performance.¹⁶ However, when assessing the calibration factor for the current study period (2002–5), we have seen that the logistic EuroSCORE requires a calibration factor of 0.53, which shows that there are ongoing improvements in the quality of surgical outcomes in the north west of England.²⁸ These differences are small, but lead to marked differences in the patients defining as exhibiting outlying performance at either 95% or 99% control limits, which have potentially important implications for public reporting of outcomes or use in professional revalidation and emphasise the need to ensure risk models are appropriately calibrated at all times to ensure accurate contemporary peer-group benchmarking.

The benefits of the locally derived model are that it is built for purpose. The advantage of the EuroSCORE is that it is widely used throughout the UK and beyond, and the tools necessary to derive its score are easily available.¹⁹ This study has shown that the EuroSCORE is a good risk adjustment model as long as it is accurately calibrated for contemporary practice but it also demonstrates that a model built for purpose would have advantages over the calibrated EuroSCORE.

If the facilities to create a local or national risk-adjustment model built for purpose were available then the application of such a model would provide a more accurate risk-adjustment model than the calibrated EuroSCORE and may lead to different conclusions about surgeons with high mortality outcomes.

Unanswered questions and future research

This study has shown clearly that crude mortality analysis or those involving the non-calibrated EuroSCORE may be misleading. All the analyses we have performed have been based on mortality, and this is only present in a small proportion of cases. It is likely that analysis of other outcomes such as stroke rate, re-exploration for bleeding, blood and blood product use, hospital stay, critical care stay and the incidence of hospital-acquired infection would be useful to enable surgeons to benchmark their outcomes against their peers and stimulate further improvements in quality, and it is probable that these analyses would also require similar risk-adjusted treatments to the ones we have shown.

Acknowledgements: This study has been conducted on behalf of the North West Quality Improvement Programme in Cardiac Interventions, and the consultant surgeons involved are listed as follows: John Au, Ben Bridgewater, Colin Campbell, John Carey, John Chalmers, Walid Dhimis, Abdul Deiraniya, Andrew Duncan, Brian Fabri, Elaine Griffiths, Geir Grotte, Ragheb Hasan, Tim Hooper, Mark Jones, Daniel Keenan, Neeraj Mediratta, Russell Millner, Nick Odom, Brian Prendergast, Mark Pullan, Abbas Rashid, Franco Sogliani, Paul Waterworth, Nizar Yonan. The following surgeons have left the collaboration during the study period: Narinda Bhatnagar, Albert Fagan, Bob Lawson, Udin Nkere. Peter O'Keefe, Richard Page, Ian Weir and David Sharpe.

Funding: Funding for the North West Quality Improvement Programme in Cardiac Interventions collaboration has been received from all primary care trusts in the north west of England. All authors were independent of the funding.

Competing interests: BB is a Society of Cardiothoracic Surgeons of GB and Ireland representative on the joint Society of Cardiothoracic Surgeons, Healthcare commission, Department of Health group defining national cardiac surgical audit and is on the executive committee of the SCTs. BB, AG, GJJ, BF and MJ are all members of the steering group of the Northwest Regional Quality Improvement Programme in Cardiac Interventions.

REFERENCES

1. *Guardian*. <http://society.guardian.co.uk/nhsperformance/story/0,,1439210,00.html>. Accessed 21 Sep 2005.
2. **Bridgewater B**, on behalf of the adult cardiac surgeons of NW England. Named surgeon mortality data for adult cardiac surgery: results on 10 157 patients undergoing coronary artery surgery and aortic valve replacement under 25 consultant surgeons in northwest England. *BMJ* 2005;**330**:506–10.
3. **Healthcare Commission**. <http://heartsurgery.healthcarecommission.org.uk>. Accessed 8 Oct 2006.
4. **Chassin MR**, Hannan EL, DeBuono BA. Benefits and hazards of reporting medical outcomes publicly. *N Engl J Med* 1996;**334**:394–8.
5. **Keogh BE**, Kinsman R. *Fifth national adult cardiac surgical database report 2003*. London: Society of Cardiothoracic Surgeons of Great Britain and Ireland, 2003.
6. **Burack JH**, Impellizzeri P, Homel P, et al. Public reporting of surgical mortality: a survey of New York State cardiothoracic surgeons. *Ann Thorac Surg* 1999;**68**:1195–2000.
7. **Narins CR**, Dosier AM, Ling FS, et al. The Influence of public reporting of outcome data on medical decision making by physicians. *Arch Intern Med* 2005;**165**:83–7.
8. **Schneider E**, Epstein A. Influence of cardiac surgery performance of referral practices and access to care. *N Engl J Med* 1996;**335**:251–6.
9. **Turi TG**. The big chill: the deleterious effects of public reporting on access to health care for the sickest patients. *J Am Coll Cardiol* 2005;**45**:1766–8.
10. **Werner RM**, Asch DA. The unintended consequences of publicly reporting quality information. *JAMA* 2005;**293**:1239–44.
11. **Bridgewater B**, Grayson AD, Jackson M, et al. Surgeon specific mortality in adult cardiac surgery: comparison between crude and risk stratified data. *BMJ* 2003;**327**:13–7.
12. **Roques F**, Nashef SA, Michel P, et al. Risk factors and outcome in European cardiac surgery: analysis of the EuroSCORE multinational database of 19,030 patients. *Eur J Cardiothorac Surg* 1999;**15**:816–23.
13. **Roques F**, Michel P, Goldstone AR, et al. The logistic EuroSCORE. *Eur Heart J* 2003;**24**:1–2.
14. **Gogbashian A**, Sedrayan A, Treasure T. EuroSCORE: a systematic review of international performance. *Eur J Cardiothorac Surg* 2004;**25**:695–700.
15. **Karthik S**, Srinivasan AK, Grayson AD, et al. Limitations of additive EuroSCORE for measuring risk stratified mortality in combined coronary and valve surgery. *Eur J Cardiothorac Surg* 2004;**26**:318–22.
16. **Bhatti F**, Grayson AD, Grotte G, et al. The logistic EuroSCORE in cardiac surgery: how well does it predict operative risk? *Heart* 2006;**92**:1817–20.
17. **Department of Health**. *Trust assurance and safety: the regulation of health professionals*. White paper. www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_065946.
18. www.nwheartaudit.nhs.uk. Accessed 8 Oct 2006.
19. www.euroscore.org. Accessed 8 Oct 2006.
20. **Hosmer D**, Lemeshow S. *Applied logistic regression*. New York: John Wiley, 1989.
21. **Hanley JA**, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982;**143**:29–36.
22. **Spiegelhalter D**. Funnel plots for comparing institutional performance. *Stat Med* 2005;**24**:1185–202.
23. **Hannan EL**, O'Donnell JF, Kilburn H Jr, et al. Investigation of the relationship between volume and mortality for surgical procedures performed in New York State hospitals. *JAMA* 1989;**262**:503–10.
24. **Parsonnet V**, Dean D, Bernstein AD. A method of uniform stratification of risk for evaluating the results of surgery in acquired heart disease. *Circulation* 1989;**79**:13–12.
25. **Wynn-Jones K**, Jackson M, Grotte GJ, et al. Limitations of the Parsonnet score for measuring risk stratified mortality in the north west of England. The North West Regional Cardiac Surgery Audit Steering Group. *Heart* 2000;**84**:71–8.
26. **Michel P**, Roques F, Nashef SA, EuroSCORE Project Group. Logistic or additive EuroSCORE for high-risk patients? *Eur J Cardiothorac Surg* 2003;**23**:684–7.
27. **Yap CH**, Mohajeri M, Ihle BJ, et al. Validation of Euroscore model in an Australian patient population. *Aust NZ J Surg* 2005;**75**:508–12.
28. **Bridgewater B**, Grayson AD, Brooks N, et al. Has the publication of cardiac surgery outcome data been associated with changes in practice in northwest England? An analysis of 25 730 patients undergoing CABG surgery under 30 surgeons over eight years. *Heart* 2007;**93**:744–8.