

Predicting operative risk for coronary artery surgery in the United Kingdom: a comparison of various risk prediction algorithms

B Bridgewater, H Neve, N Moat, T Hooper, M Jones

Abstract

Objective—To compare the ability of four risk models to predict operative mortality after coronary artery bypass graft surgery (CABG) in the United Kingdom.

Design—Prospective study.

Setting—Two cardiothoracic centres in the United Kingdom.

Subjects—1774 patients having CABG.

Main outcome measures—Risk factors were recorded for all patients, along with in-hospital mortality. Predicted mortality was derived from the American Society of Thoracic Surgeons (STS) risk program, Ontario Province risk score (PACCN), Parsonnet score, and the UK Society of Cardiothoracic Surgeons risk algorithm.

Results—There were significant differences ($p < 0.05$) between the British and American populations from which the STS risk algorithm was derived with respect to most variables. The observed mortality in the British population was 3.7% (65 of 1774). The mean predicted mortality by STS score, PACCN, Parsonnet score, and UK algorithms were 1.1%, 1.6%, 4.6%, and 4.7% respectively. The overall predictive ability of the models as measured by the area under the receiver operating characteristic curve were 0.64, 0.60, 0.73, and 0.75, respectively.

Conclusions—There are differences between the British and American populations for CABG and the North American algorithms are not useful for predicting mortality in the United Kingdom. The UK Society of Cardiothoracic Surgeons algorithm is the best of the models tested but still only has limited predictive ability. Great care must be exercised when using methods of this type for comparisons of units and surgeons.

(Heart 1998;79:350-355)

Keywords: cardiac surgery; risk stratification; audit

In the modern era, results of medical care given by different institutions and doctors are increasingly scrutinised, and the production of hospital league tables is now widespread. Such league tables usually fail to account for case mix and severity, and great caution should be exercised in their interpretation. The development of tools to allow results from different hospitals and surgeons to be compared in a

meaningful way is obviously an important goal for internal and external audit.

Operative mortality is an indicator of the quality of cardiac surgery. Comparing different institutions or surgeons on the basis of crude mortality figures may be misleading as mortality is affected by various preoperative characteristics of the patients,¹ and before comparisons can be made it is important to adjust the mortality by accounting for these characteristics. The ideal risk stratification model should be easy to implement, objective, an accurate predictor of observed mortality, and in widespread use.

Various models have been developed for use in cardiac surgery, and the first to become popular was the Parsonnet risk stratification system² which was developed in the USA in the 1980s. The model allocates additive predicted mortality percentage points to 14 patient risk factors to give a "Parsonnet score," which is supposed to be indicative of the percent mortality for each patient and has been used to categorise patients into good risk (predicted mortality 0% to 4%), fair risk (5% to 9%), poor risk (10% to 14%), high risk (15% to 19%), and extremely high risk (> 20%). It has been shown to be applicable to British cardiac surgical practice.³ However, it has been criticised for the nature of its statistical derivation,^{4,5} it systematically overestimates mortality, particularly for high risk patients, and its scoring system is quite subjective, again especially in the high risk group. It also omits many surgeons' "favourite" risk factors, such as the number of coronary vessels diseased, urgency of operation, and presence of chronic obstructive pulmonary disease.⁶

More recently other models have been developed. The Society of Thoracic Surgeons (STS) in the USA has produced an algorithm to predict operative mortality for patients undergoing coronary artery surgery alone.⁷ It is based on a very large patient population and has been developed using the Bayes theorem. It has been shown to be a good predictor of observed mortality in the USA.^{7,8} In Canada, a risk index has been developed to predict mortality, prolonged stay in the intensive care unit, and prolonged hospital stay⁹ in all patients undergoing cardiac surgery. In the United Kingdom a model has recently been produced to predict mortality following coronary artery bypass graft surgery (CABG).¹⁰ We have studied the ability of these models to predict mortality after CABG in the United Kingdom.

Departments of
Cardiothoracic
Surgery and Clinical
Audit, Wythenshawe
Hospital, Manchester,
UK

B J M Bridgewater
H J Neve
T L Hooper
M T Jones

Department of
Cardiac Surgery,
Royal Brompton
Hospital, London SW3,
UK
N Moat

Correspondence to:
Mr B J M Bridgewater,
Department of
Cardiothoracic Surgery,
Green Lane Hospital, Green
Lane West, Auckland 3, New
Zealand.
email: bbridge@nznet.gen.nz

Accepted for publication
3 October 1997

Table 1 Definition of risk factors

Risk factor	Definition
Smoking history	Any history of tobacco use
Diabetes	Any history of diabetes regardless of duration
Renal failure	Blood creatinine > 140 µmol/l or on dialysis
Hypertension	Blood pressure > 140/90 mm Hg or on antihypertensive drugs
Cerebrovascular accident	Central neurological deficit persisting > 24 hours or previous cerebrovascular surgery
Chronic obstructive pulmonary disease	Pharmacological therapy for COPD or FEV ₁ < 75% predicted
Unstable angina	Myocardial ischaemia requiring hospitalisation and use of iv medication for control
Urgent surgery	Surgery within 24 hours of referral
Emergency surgery	No delay in providing operative intervention
Salvage	Cardiopulmonary resuscitation on route to operating theatre or during anaesthesia

Definitions as given in the Society of Thoracic Surgeons National Cardiac Surgery database manual for data managers.¹³

Methods

We have studied all patients undergoing CABG alone in two centres: Wythenshawe Hospital, Manchester, and The Royal Brompton Hospital, London.

WYTHENSHAW HOSPITAL

The patients at Wythenshawe Hospital were recruited between February 1995 and January 1996. Risk factors were recorded in a computerised database by the operating surgeon at the time of surgery, from data collected on admission on structured patient clerking sheets. Data were checked for completeness and omitted or erroneous data were completed by reference to hospital notes. Once a complete dataset was obtained for each patient there was no further verification.

ROYAL BROMPTON HOSPITAL

The patients from the Royal Brompton Hospital were those operated upon between April 1994 and March 1995. Data were collected on structured clerking sheets by junior medical staff and transcribed onto a computerised data base by clerical staff. Once again, there was no subsequent verification once a complete dataset had been obtained.

PATIENT RISK FACTORS

The definitions used for some of the risk factors are shown in table 1. These definitions are those given by the Society of Thoracic Surgeons for the USA national database.¹¹ Renal failure at Wythenshawe Hospital was defined as plasma creatinine greater than 140 µmol/l or dialysis dependent, and at The Royal Brompton Hospital as plasma creatinine greater than 150 µmol/l, urea greater than 15 mmol/l, or requiring dialysis. For ejection fraction a single estimated figure was collected at Wythenshawe, and at the Royal Brompton it was graded as good (50% or more), fair (30 to 49%), or poor (less than 30%).

RISK SCORES

The Parsonnet score was derived from the risk factors as described previously and shown in table 2.² For the purposes of this study we have automatically given an additional score of 10 points for a catastrophic state, rather than allowing between 10 and 50 as suggested in the

original Parsonnet system, to decrease subjectivity on the higher risk patients. To calculate the STS risk score the data were imported into the commercially available STS risk stratification software (Summit Medical Systems, Nice, France) and a risk prediction for each patient was calculated. The PACCN score was derived as shown in table 3,⁹ and the UK Society of Cardiothoracic Surgeons score (UK national score), as published using the Bayes theorem, is shown in table 4.¹⁰

OPERATIVE MORTALITY

Mortality was recorded from patient and hospital records. Operative mortality was defined as death temporally or causally related to surgery (death within 30 days of operation or in the same hospital admission as operation, regardless of cause). The cause of death was recorded from patient records or necropsy examination when available.

Table 2 Parsonnet score

Risk factor	Score*
Female	1
Obesity (> 1.5 times ideal weight)	3
Diabetes	3
Hypertension	3
Ejection fraction (%)	
> 50	0
30-49	2
< 30	4
Age (years)	
70-74	7
75-80	12
≥ 80	20
First reoperation	5
Second reoperation	10
Preoperative IABP	2
LV aneurysm (resected)	5
Emergency from PTCA/ cath lab	10
Dialysis dependent	10
Catastrophic states	10-50
Rare circumstances	2-10
Valve surgery	
Mitral	5
PA > 60 mm Hg	8
Aortic	5
Gradient > 120 mm Hg	7
CABG with valve	2

*Parsonnet scoring system.²

IABP, intra-aortic balloon pump; PA, pulmonary artery systolic pressure.

Table 3 The PACCN score

Risk factor	Score*
Age (years)	
< 65	0
65-74	2
≥ 75	3
Sex	
Male	0
Female	1
Left ventricular ejection fraction	
> 50%	0
35-50%	1
20-34%	2
< 20%	
Type of surgery	
CABG only	0
Single valve	2
Complex	3
Urgency of surgery	
Elective	0
Urgent	1
Emergency	4
Repeat operation	
No	0
Yes	2
Range of scores	0 to 16

*PACCN score.⁹

Table 4 The UK society score¹⁰

Preoperative factor		Death (%)	Likelihood ratio	Weight of evidence
Overall		5.2	0.05	-29
Age (years)	20 to 39	1.5	0.4	-9
	40 to 49	2.3	0.4	-8
	50 to 59	3.5	0.7	-4
	60 to 69	5	1	0
	70 to 79	9.2	1.9	6
	80 to 88	22.2	5.4	17
Sex	Male	5.1	1	0
	Female	8.4	1.7	5
Angina	None to class IV	2.7 to 7.7	0.5 to 1.5	-7 to 4
Dyspnoea	None to class IV	3.8 to 14.8	0.7 to 3.2	-1 to 12
COPD	No	5.7	1.1	1
	Yes	9.4	2	7
Emphysema	No	5.8	1.1	1
	Yes	7.7	1.6	5
Asthma	No	5.8	1.1	1
	Yes	7.1	1.4	4
Pre-op support	None	5.2	1	0
	IABP	28.6	8.3	21
	Ventilated	25	7.8	21
	Inotropes	17.6	4.4	15
	> 1 of the above	58.3	24.8	32
Renal	None	4.1	0.8	-2
	History	12.3	2.6	10
	Dialysis	12.5	3	11
Smoking	Never	5.6	5.6	1
	Stopped < 5 years	3.6	0.7	-4
	Stopped > 5 years	5.7	1.1	1
	Smoker	6.3	1.2	2
Diabetes	None	5	1	0
	Diet	7.6	1.6	4
	Oral medication	5.7	1.1	1
	Insulin	6.6	1.4	-2
Hypertension	No	4.3	0.8	-2
	Yes	6.4	1.3	2
Vascular disease	No	4.1	0.8	-2
	Yes	8.4	1.7	5
Cerebral	No	5	1	0
	TIA	12.4	2.7	10
	CVA	10.5	2.2	8
LMS	No	5	1	0
	Yes	6.4	1.3	2
Last MI	None	4.7	0.9	-1
	< 24 hours	22.7	5.5	17
	1 to 30 days	9.1	1.9	6
	> 30 days	5.2	1	0
Last PTCA	None	5	1	0
	< 24 hours	10.3	2.2	8
	1 to 30 days	5.6	1.3	3
	> 30 days	4.6	0.9	-1
Ejection fraction	0 to 29%	7.1	1.4	4
	30 to 50%	4.6	0.9	-1
	> 50%	4.9	0.9	-1
Op sequence	First	4.7	0.9	-1
	Re-operation	13.3	2.8	10
Op timing	Elective	3.5	0.7	-4
	Urgent	6.4	1.2	2
	Emergency	14.1	3	11
	Salvage	40	12.4	25

LMS, left main stem stenosis > 50%; COPD, chronic obstructive pulmonary disease.

Bayes theorem is the formula that provides the relation between the prior odds and posterior odds: posterior odds = likelihood ratio × prior odds; where the likelihood ratio expresses how much more likely it is that a patient with a given condition should fall among those who die rather than survive. Because this is repeated for each risk factor assuming they are contributing separately, it usually leads to an overprediction of mortality and the contribution of each factor should be damped down accordingly.

For this study we have used: posterior odds = likelihood ratio^{0.8} × prior odds.

To turn the risk factor table into a risk prediction a weight of evidence is calculated that expresses how much an observed feature adds to or subtracts from the evidence of mortality where weight of evidence = 10 log (likelihood ratio).

The scoring system therefore consists of adding up the weights of evidence for the observed features and adding the results to the starting score of -29, which is the 10 log (number of deaths/number of survivors). This gives a total score "S" that may be transformed to a predicted percentage mortality by the equation: $p = 100/[1 + \exp(-S/10)]$.

DATA ANALYSIS

The incidence of risk factors between the two centres in the United Kingdom, and between the combined Wythenshawe/Brompton population and those reported by the STS in the USA from which the STS risk prediction algorithm was generated, were compared by χ^2 test. Ninety five per cent confidence limits for the

British and American populations are also given, along with the relative risk (RR). The observed mortality in our sample was compared with the overall mortality in the United Kingdom and the USA, again by χ^2 test. The distributions of the risk predictions were not normally distributed and so the mean scores, median scores, and ranges are given. Receiver operating characteristic (ROC) curves were plotted for the predictive value of the different models and the area under the ROC curves was calculated as an index of the overall predictive ability of the models.

Results

PREOPERATIVE PATIENT CHARACTERISTICS

The incidence of risk factors at Wythenshawe Hospital and the Royal Brompton Hospital is shown in table 5, and those for the pooled United Kingdom population and those reported in the USA by the STS are shown in table 6. The average age of the British population was 60.6 years (range 32 to 83 years); 82% were male and 18% female. There were significant differences between the two British centres with respect to age greater than 70 years, smoking history, renal failure, chronic obstructive pulmonary disease, previous cerebrovascular accident, unstable angina, and redo surgery. There were significant differences between the British and the American populations with respect to all variables except renal failure and redo surgery. Morbid obesity (relative risk (RR) 7.2), ejection fraction less than 50% (RR 3), unstable angina (RR 1.9), age greater than 70 years (RR 1.8), diabetes (RR 1.5), female sex (RR 1.4), hypertension (RR 1.3), and non-elective surgery (RR 1.2) were more prevalent in the USA. Intravenous nitrates (RR 0.1), previous cerebrovascular accident (RR 0.4), left main stem disease (RR 0.5), smoking history (RR 0.6), recent percutaneous transluminal coronary angioplasty (PTCA) (RR 0.7), and chronic obstructive pulmonary disease (RR 0.7) were more common in the United Kingdom.

OPERATIVE MORTALITY

Documentation of operative mortality was exhaustive and we have obtained alive/dead status on all patients. The operative mortality was 65 of 1774 patients (3.7%), 39 of 995 at Wythenshawe (3.9%) and 26 of 779 at the Royal Brompton (3.3%). This difference was not significant. The overall mortality was similar to the 3.5% reported in the United Kingdom cardiac surgical register 1994/95 and slightly higher than 2.8% for the STS 1991-1993 ($p < 0.05$). The commonest causes of death were cardiac in 68%, respiratory in 15%, and sepsis 5%. The causes of death in the British and American populations are shown in table 7. Respiratory deaths were more common in the British population ($p < 0.05$), and multisystem failure more common in the USA ($p < 0.05$).

RISK PREDICTIONS

The mean overall predicted mortality was 1.07% (median 0.9, range 0.3% to 11.5%) by

Table 5 Percentage incidence in risk factors at the two UK centres

Risk factor	Wythenshawe, n = 995 (95% CL)	Brompton, n = 779 (95% CL)	p value
Mortality	3.3 (2.4–4.7)	3.9 (2.6–5.30)	NS
Age > 70 years	10.2 (8.5–12.3)	16.2 (13.8–19.0)	< 0.001
Female	19.1 (16.8–21.7)	16.9 (14.5–19.7)	NS
Morbid obesity	3.7 (2.7–5.1)	2.6 (1.8–4.0)	NS
Hypertension	40.5 (37.5–43.6)	43.8 (40.0–47.4)	NS
Diabetes	11.4 (9.6–13.6)	12.7 (10.6–15.3)	NS
COPD	6 (4.7–7.7)	2.4 (1.6–3.5)	< 0.001
Renal failure	3.7 (2.7–5.1)	1.4 (0.8–2.6)	< 0.001
EF < 50%	18.1 (15.9–20.7)	17.6 (15.1–20.5)	NS
Redo	4.1 (3.1–5.6)	7.6 (6.0–9.7)	< 0.01
Non-elective	16.2 (14.1–18.7)	17.3 (15.9–20.2)	NS
Unstable angina	19.7 (17.4–22.3)	34.8 (31.6–38.3)	< 0.01
LMS disease	21.3 (18.9–24.0)	18.2 (15.7–21.1)	NS
Recent PTCA	3.3 (2.4–4.7)	1.8 (1.1–3.1)	NS
CVA	4.3 (3.2–5.8)	3 (2.1–4.5)	< 0.05
Smoking history	72.5 (70–75.2)	57.1 (53.6–50.1)	< 0.01
Intravenous nitrates	8.7 (7.1–10.7)	9.8 (8.0–12.2)	NS

CL, confidence limits; COPD, chronic obstructive pulmonary disease; CVA, cerebrovascular accident; EF, ejection fraction; LMS, left main stem; PTCA, percutaneous transluminal coronary angioplasty.

Table 6 Percentage incidence of risk factors in UK and USA practice

Risk factor	UK (n = 1774)	USA (n = 78,927)	p value	Relative risk (UK v USA)
Mortality	3.7 (2.9–4.7)	3.2 (3.1–3.3)	NS	0.86
Age > 70 years	12.8 (11.3–14.5)	22.6 (22.3–22.9)	< 0.001	1.77
Female	18.2 (16.5–20.1)	24.6 (24.3–24.9)	< 0.001	1.35
Morbid obesity	3.2 (2.5–4.2)	23 (22.7–23.3)	p < 0.001	7.18
Hypertension	41.9 (39.6–44.2)	53.4 (53.0–53.7)	p < 0.001	1.27
Diabetes	12.1 (10.7–13.7)	17.9 (17.6–18.2)	p < 0.001	1.48
COPD	4.5 (3.6–5.6)	3.2 (3.1–3.3)	p < 0.01	0.71
Renal failure	2.7 (2.1–3.6)	2.2 (2.1–2.3)	NS	0.81
EF < 50%	17.9 (16.2–19.8)	52.9 (52.6–53.2)	p < 0.001	3.00
Redo	5.6 (4.6–6.7)	6.7 (6.5–6.9)	NS	1.20
Non-elective	16.7 (15.0–18.5)	20.1 (19.8–20.4)	p < 0.001	1.20
Unstable angina	21.9 (20.1–23.9)	41 (40.7–41.3)	p < 0.001	1.87
LMS disease	20 (18.2–21.9)	9.3 (9.1–9.5)	p < 0.001	0.46
Recent PTCA	2.6 (2.0–3.5)	1.8 (1.7–1.9)	p < 0.05	0.69
CVA	3.7 (2.9–4.7)	1.4 (1.3–1.5)	p < 0.001	0.37
Smoking history	65.7 (63.5–67.9)	42.7 (42.4–43.0)	p < 0.001	0.65
Intravenous nitrates	9.2 (8.0–10.7)	0.85 (0.79–0.92)	p < 0.001	0.09

COPD, chronic obstructive pulmonary disease; CVA, cerebrovascular accident; EF, ejection fraction; LMS, left main stem; PTCA, percutaneous transluminal coronary angioplasty.

the STS algorithm, 1.6% (median 1.5, range 0% to 9%) by the PACCN score, 4.6% (median 3, range 0% to 45%) by Parsonnet score, and 4.7% (median 4.3%, range 0.5% to 76%) by the UK national score.

ROC CURVES

The area under the ROC curves were 0.64 for the STS algorithm, 0.60 for the PACCN score, 0.73 for the Parsonnet score, and 0.75 for the UK national algorithm. An area of 1 suggests a perfect predictor, and a value of 0.5 is a test of no value. Areas of between 0.5 and 0.7 represent a low accuracy, between 0.7 and 0.9 are useful for some purposes, and higher values represent a high accuracy.¹²

Table 7 Causes of death

Cause of death	UK (n = 1774) (% of total deaths)	USA (n = 78927) (% of total deaths)	p value
Myocardial	67.7	61	NS
Respiratory	15.4	7.7	p < 0.05
Neurological	1.5	7.7	NS
Sepsis	4.6	4.6	NS
Multisystem failure	1.5	15.7	p < 0.05
Renal	3.1	3.3	NS
Other	7.9	15.7	NS

Discussion

Developing an appropriate tool for predicting risk is an important goal, because only then can operative results be viewed in the context of case mix. Risk stratified mortality data are useful for internal audit processes to allow surgeons to review their results compared with their peers to ensure that the quality of the service they provide is satisfactory. It also allows external audit to be performed in a meaningful way, enabling units to be compared by purchasers or other interested parties. In the increasingly aggressive medicolegal environment, accurate collection of this type of data analysed in a responsible way also provides protection for clinicians.

Before a risk prediction algorithm is used for these aims, the dataset to which it is applied should be validated and trusted, the risk prediction tool should be shown to be an accurate predictor of operative mortality, the system should only be used on the types of practice on which it has been validated, comparisons between populations should not be drawn if there are great differences in case mix, and conclusions should only be drawn after an intelligent and responsible analysis with an informed knowledge of all the factors.

The data presented here have been collected from two surgical units, one in the southeast and one in the northwest of England. Together we perform about 2800 open heart procedures each year and each department supports a busy interventional cardiology programme. We have no reason to think that our case mix is atypical of busy regional units and our observed mortality for coronary artery surgery in the 1774 patients reported here is 3.7%, which is similar to that reported in the United Kingdom Cardiac Surgical Register in 1994/95.¹³ There was no difference in the overall mortality between the two units. All of the patients in the period studied have been risk stratified, and documentation of 30 day mortality has been exhaustive and complete.

There were some differences in the incidence of patient risk factors between the two centres used in this study: the incidence of age greater than 70, redo surgery, and unstable angina was higher at the Brompton Hospital, but chronic obstructive pulmonary disease, current smoking history, and renal failure were more common at Wythenshawe Hospital, reflecting the nature of the referral patterns and the socioeconomic makeup of the different regions.

There are notable differences between the pooled British patients and those in the USA from which the STS model was derived. Some of these differences are conflicting: for example, the incidence of non-elective surgery was higher in the USA (20% v 16%), but the incidence of intravenous nitrate use was higher in the United Kingdom (1% v 9%). This paradox is a result of the definitions used and the different nature of health care provision in the two systems. The definitions of salvage, emergency, urgent, and elective operations shown in table 1 were developed for American practice and for the purposes of this study we have used them here. Because of the pressures on the NHS, we

almost never operate within 24 hours of referral, unless it is an “emergency” operation, and so a patient who has coronary angiography and is found to have a tight left main stem stenosis and a blocked right coronary artery may well receive an operation the following day in the USA and so be an urgent case, when they would probably wait for several days in our hospitals and so be an elective case according to the definitions used. This would cause an underestimation of the predicted mortality according to the STS model in the United Kingdom.

The definitions of some of the risk factors are objective and easily compared, such as renal failure, but others are more difficult. For example chronic obstructive pulmonary disease is defined as any case where drugs are given for chronic obstructive lung disease or where the FEV₁ (forced expiratory volume in one second) is less than 75% of the predicted value. This seems to be an important risk factor in our practice, because 15% of our mortality was due to respiratory causes, compared to 7% in the USA. According to this definition, the incidence of chronic obstructive pulmonary disease in the United Kingdom is 4.5%, against 3.2% in the USA, but we do not measure lung function preoperatively and so only pick up those patients with a history or who are on drug treatment. This underdetection of the incidence of chronic obstructive pulmonary disease leads to an underprediction of risk and may be another reason why the STS model significantly underestimates our observed mortality. Some of the differences between the United Kingdom and the USA are statistically large but clinically small (for example, non-elective operations: 16.7% in the United Kingdom, 20.1% in the USA, RR 1.2) and are driven by the huge numbers in the groups, but they are important as the incidence of the risk factors in the population and their associations with mortality in that group are the way in which the risk models are developed. Any difference in the populations casts a doubt on the wisdom of using a risk model produced from one population for predicting risk in another statistically different population. Other risk factors also show a marked difference between the two populations (for example, morbid obesity 3.2% in the United Kingdom, 23% in the USA, RR 7.2)

Assessing the accuracy of a multivariate prediction tool is not straightforward. The predictive equations provide a probability of mortality of between 0% and 100% for each patient, based on the incidence of risk factors, but each patient either dies or does not and so direct comparisons of predicted and observed outcome for the individual patient are not a useful measure of the efficacy of the predictive tool. The area under the ROC curve is believed to be a more appropriate statistical measure of the ability of a model to predict what it intends to. It is a plot of sensitivity versus 1 – specificity, and the area under the curve is a useful summary measure of the diagnostic accuracy of the tool. An area of 1 suggests a perfect predictor, and a value of 0.5 is a test of no value.

Areas of between 0.5 and around 0.7 represent a rather low accuracy—the true positive proportion is not much greater than the false positive proportion. Values between 0.7 and 0.9 are useful for some purposes, and higher values represent a high accuracy.^{12–16} ROC curves have now been widely used for evaluating the accuracy of risk prediction models.^{9–14–17–18}

The overall predictive ability of the STS model as indicated by the area under the ROC curve is 0.64, indicating it is not useful for British practice. The PACCN score was derived from analysis of 6213 patients undergoing cardiac surgery in Ontario province, Canada, in 1991.⁹ The model was developed using multivariate analysis to determine important risk factors which were then combined into an additive model. The model was then validated on a test set of 6885 patients from the same region in 1992 and shown to predict mortality, very long intensive care stay, and very long in postoperative stay, with areas under the ROC curve of 0.75, 0.67, and 0.71 respectively. This risk index does not include any additional score for comorbid, non-cardiac disease, which may account for its poor performance in predicting mortality in our analysis. The area under the ROC curve of 6.0 was the worst of all the models tested.

As has been reported previously,³ the Parsonnet score provides a reasonable predictor of operative death for British practice, but it consistently overpredicts observed mortality, particularly for the higher risk groups. The Parsonnet scores ranged from 0% to 45%, with a mean of 4.7% and a median of 3, and while the scores are not normally distributed the mean Parsonnet score for the whole group of 4.7% gives some idea of case mix and enables the overall mortality of 3.7% to be viewed in that context. As well as overestimating risk in contemporary clinical practice, the Parsonnet score has been criticised for being subjective at higher levels of risk. We have tried to minimise this in this study by simply including an additional score of 10 points for any patient in a catastrophic state, rather than allowing a range of 10 to 50 points as suggested in the original model.² The ROC curve analysis suggests that the Parsonnet score has useful but limited predictive ability in this population.

The UK national score¹⁰ is still being developed but was derived from a population of 4159 patients undergoing coronary artery surgery alone from three centres. The risk data included in the study have not yet been validated fully. The model was developed using the Bayes theorem, as shown in table 4. This algorithm has not yet been validated on an independent population, but the predicted mortality for our patients of 4.7% again overestimates our observed mortality of 3.7%, though the overall predictive ability as given by an area under the ROC curve of 0.75 was the best of any of the models tested.

Other models exist for predicting mortality, such as that generated in New England¹⁴ or the Veterans Affairs Cardiac Surgery Consultants' Committee,¹⁵ but we have not studied their

predictive ability as we do not routinely collect the data required for their models.

The predictive models studied have areas under the ROC curve ranging from 0.6 to 0.75. It may be impossible to produce a model with a curve area much higher than this, as some aspects of mortality will always be related to risk factors not included in the model (for example, the quality of the distal vessels) or due to chance happenings not related to preoperative patient characteristics (such as surgical error).¹⁶

Development of an appropriate risk stratification model is an important goal in cardiac surgery, and an ideal tool should be easy to implement, objective, an accurate predictor of observed mortality, and in widespread use, allowing comparison between surgeons and units to be made readily. No existing tool fulfils all of these criteria in the United Kingdom, but the Parsonnet score and preliminary version of the UK national scores seem to be the best tools available at present. Even these only have limited ability to predict observed outcome, as shown by the area under the ROC curve of 0.73 and 0.75 respectively, and great care should still be exercised when using these models.

We would like to thank Maureen Silcock and the staff of the clinical audit department at Wythenshawe for their help with this study, and Philip Kimberley for collating the data from the Royal Brompton. We are grateful to all the consultant surgeons at both institutions for allowing us to study their patients and would also like to thank Geoff Corner for his assistance with the surgical database and Brian Farragher for statistical advice. We would also like to acknowledge the help of Professor Tom Treasure in the preparation of this manuscript. The work was supported in part by a grant from the Northwest Regional Health Authority.

- 1 DuBois RW, Rogers WH, Moxley JH, *et al.* Hospital inpatients mortality: is it a predictor of quality. *N Engl J Med* 1987;317:1674-80.
- 2 Parsonnet V, Dean D, Bernstein, AD. A method of uniform stratification of risk for evaluating the results of surgery in acquired adult heart disease. *Circulation* 1989;701(suppl): 13-112.
- 3 Nashef SAM, Carey F, Silcock MM, *et al.* Risk stratification for open heart surgery; trial of the Parsonnet system in a British Hospital. *BMJ* 1992;305:1066-7.
- 4 Allen A, Forsyth AT. Risk stratification in open heart surgery [letter]. *BMJ* 1992;305:1500-1.
- 5 Spiegelhalter DJ. Risk stratification for open heart surgery [letter]. *BMJ* 1992;305:1500.
- 6 Treasure T. Risks and results of surgery. *Br Heart J* 1995;74: 11-12.
- 7 Edwards FH, Clark RE, Schwartz M. Coronary artery bypass grafting: the Society of Thoracic Surgeons National database experience. *Ann Thorac Surg* 1994;57:12-19.
- 8 Hattler BG, Madia C, Johnson C, *et al.* Risk stratification using the Society of Thoracic Surgeons program. *Ann Thorac Surg* 1994;58:1348-52.
- 9 Tu JV, Jaglal SB, Naylor D. Multicenter validation of a risk index for mortality, intensive care unit stay, and overall hospital length of stay after cardiac surgery. *Circulation* 1995;91:677-84.
- 10 Society of Cardiothoracic Surgeons of Great Britain and Ireland National Database. Preliminary report 1996. In conjunction with the MRC Biostatistics Unit Cambridge and Dendrite Clinical systems (UK) Ltd.
- 11 The Society of Thoracic Surgeons National Cardiac Surgery database manual for data managers. Minneapolis: Summit Medical Systems, August 1993.
- 12 Swets JA. Measuring the accuracy of diagnostic systems. *Science* 1988;240:1285-93.
- 13 Society of Cardiothoracic Surgeons of Great Britain and Ireland. Cardiac Surgery Register 1994/5.
- 14 O'Connor GGT, Plume SK, Olmstead EM, *et al.* Multivariate prediction of in hospital mortality associated with coronary artery bypass graft surgery. *Circulation* 1992;85: 2110-18.
- 15 Grover FL, Johnson RR, Shroyer LW, *et al.* The Veterans Affairs continuous improvement in cardiac surgery study. *Ann Thorac Surg* 1994;58:1845-51.
- 16 Clark RE. Calculating risk and outcome: the Society of Thoracic Surgeons database. *Ann Thorac Surg* 1996;62:S2-5.
- 17 Dupuis J-Y, Wynands JE. Risk-adjusted mortality to assess quality of care in cardiac surgery. *Can J Anaesthesia* 1993;40:91-7.
- 18 Marshall G, Shroyer LW, Grover FL, Hammermeister KE. Bayesian-logit model for risk assessment in coronary artery bypass grafting. *Ann Thorac Surg* 1994;57:1492-500.