

European system for cardiac operative risk evaluation (*EuroSCORE*)[☆]

S.A.M. Nashef*, F. Roques, P. Michel, E. Gauducheau, S. Lemeshow, R. Salamon,
the *EuroSCORE* study group

Papworth Hospital, Cambridge CB3 8RE, UK

Received 21 September 1998; accepted 29 March 1999

Abstract

Objective: To construct a scoring system for the prediction of early mortality in cardiac surgical patients in Europe on the basis of objective risk factors. **Methods:** The *EuroSCORE* database was divided into developmental and validation subsets. In the former, risk factors deemed to be objective, credible, obtainable and difficult to falsify were weighted on the basis of regression analysis. An additive score of predicted mortality was constructed. Its calibration and discrimination characteristics were assessed in the validation dataset. Thresholds were defined to distinguish low, moderate and high risk groups. **Results:** The developmental dataset had 13 302 patients, calibration by Hosmer Lemeshow Chi square was $(8) = 8.26$ ($P < 0.40$) and discrimination by area under ROC curve was 0.79. The validation dataset had 1479 patients, calibration Chi square $(10) = 7.5$, $P < 0.68$ and the area under the ROC curve was 0.76. The scoring system identified three groups of risk factors with their weights (additive % predicted mortality) in brackets. Patient-related factors were age over 60 (one per 5 years or part thereof), female (1), chronic pulmonary disease (1), extracardiac arteriopathy (2), neurological dysfunction (2), previous cardiac surgery (3), serum creatinine $>200 \mu\text{mol/l}$ (2), active endocarditis (3) and critical preoperative state (3). Cardiac factors were unstable angina on intravenous nitrates (2), reduced left ventricular ejection fraction (30–50%: 1, $<30\%$: 3), recent (<90 days) myocardial infarction (2) and pulmonary systolic pressure >60 mmHg (2). Operation-related factors were emergency (2), other than isolated coronary surgery (2), thoracic aorta surgery (3) and surgery for postinfarct septal rupture (4). The scoring system was then applied to three risk groups. The low risk group (*EuroSCORE* 1–2) had 4529 patients with 36 deaths (0.8%), 95% confidence limits for observed mortality (0.56–1.10) and for expected mortality (1.27–1.29). The medium risk group (*EuroSCORE* 3–5) had 5977 patients with 182 deaths (3%), observed mortality (2.62–3.51), predicted (2.90–2.94). The high risk group (*EuroSCORE* 6 plus) had 4293 patients with 480 deaths (11.2%) observed mortality (10.25–12.16), predicted (10.93–11.54). Overall, there were 698 deaths in 14 799 patients (4.7%), observed mortality (4.37–5.06), predicted (4.72–4.95). **Conclusion:** *EuroSCORE* is a simple, objective and up-to-date system for assessing heart surgery, soundly based on one of the largest, most complete and accurate databases in European cardiac surgical history. We recommend its widespread use. © 1999 Elsevier Science B.V. All rights reserved.

Keywords: Cardiac surgery; Risk stratification; Mortality

1. Introduction

The purpose of this work was to use the *EuroSCORE* project database to construct a risk stratification system to help in the assessment of the quality of cardiac surgical care.

2. Methods

The *EuroSCORE* project set-up, data collection and entry

[☆] Presented at the 12th Annual Meeting of the European Association for Cardio-thoracic Surgery, Brussels, Belgium, September 20–23, 1998.

* Corresponding author. Tel.: +44-1480-830541; *EuroSCORE* website: euroscore.co.uk.

E-mail address: sam.nashef@papworth-tr.anglox.nhs.uk (S.A.M. Nashef)

and quality checks have been described elsewhere and multiple regression analysis had already identified a number of risk factors associated with postoperative mortality [1]. These factors were then evaluated by an international panel of cardiac surgeons with an interest in risk stratification in the hope of identifying those risk factors most likely to be useful in a risk model. The evaluation was on the basis of objectivity, credibility, availability and resistance to falsification. Factors deemed to satisfy these criteria were used for the construction of the model.

The database was randomly divided into two subsets: a developmental dataset which served for the construction of the risk model, and a validation subset for testing and validating the model. In the developmental subset, variables entered in the model were selected using bivariate tests, chi square tests for categorical covariates and *t*-tests or

Table 1
SCORE datasets

Calibration model Dataset	Patients	Discrimination model Chi square (Hosmer Lemeshow)	Area under ROC curve
Developmental	13302	Chi2(8) = 8.26, P < 0.40	0.79
Validation	1497	Chi2(10) = 7.5, P < 0.68	0.76

Wilcoxon rank sum tests for continuous covariates. All variables significant at the $P < 0.2$ level were entered into the model provided they were present in at least 2% of the sample. Non-significant variables were eliminated from the model one at a time, beginning with the variable having the highest P -value. Stability of the model was checked every time a variable was eliminated. In the case of continuous variables where the relationship with outcome was not linear, such as age and serum creatinine, we determined cut-off points using the fractional polynomials method. When all statistically non-significant variables had been eliminated from the model, goodness-of-fit testing (Hosmer Lemeshow Chi square) was used to assess how well the model was calibrated and the area under the receiver operating characteristic (ROC) curve was used to assess how well the model could discriminate between patients who lived and patients who died.

After initial assessment of the performance of the model, variables whose elimination improved calibration while not significantly affecting discrimination were dropped from the model one at a time. Although significantly associated with outcome, urgent operation and chronic congestive heart failure were eliminated because they were liable to distortion. The final step was to search for first-degree interaction. The criteria for including an interaction term were that it had to be significant at $P < 0.05$, 1% of the sample had to exhibit that combination of factors and the combination had to be clinically relevant.

The weights attributed to each variable in the score were obtained from the logistic regression b-coefficients. The calibration and discrimination power of the model were then assessed in the validation dataset. The scoring system

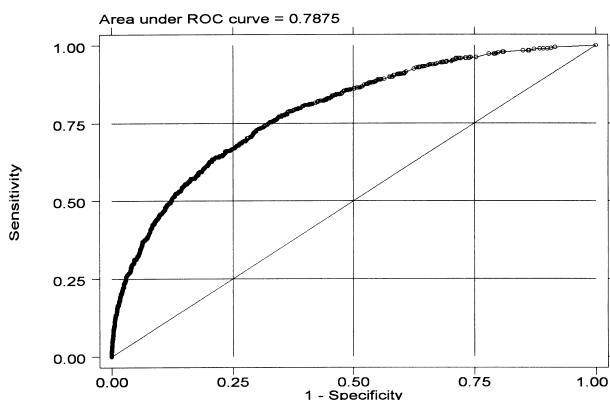


Fig. 1. ROC curve graphs for the developmental dataset ($n = 13302$).

was then used to define three risk groups (low, medium and high risk). The thresholds were chosen so that the groups would be of similar size.

3. Results

The statistical features of the developmental and validation datasets are in Table 1 and Figs. 1 and 2. The validation analysis confirmed that the model performed well both in its calibration and discrimination characteristics. Seventeen risk factors were weighted for the definitive scoring system. There were nine patient-related factors, four factors were derived from the preoperative cardiac status and four depended on the timing and nature of the operation performed. The risk factors, their definitions and the weights allocated to them are detailed in Table 2. The system is additive: to calculate the predicted risk for a patient, the scores for existing risk factors are added to give an approximate percentage predicted mortality figure. When the scoring system was applied in three different risk groups, there was very good overlap between observed and expected mortality in all three groups (Table 3).

4. Discussion

Those who provide, purchase and use health care recognise that the resources for such care are limited. It is now established that the cost of treatment must be taken into consideration in decisions about health care provision.

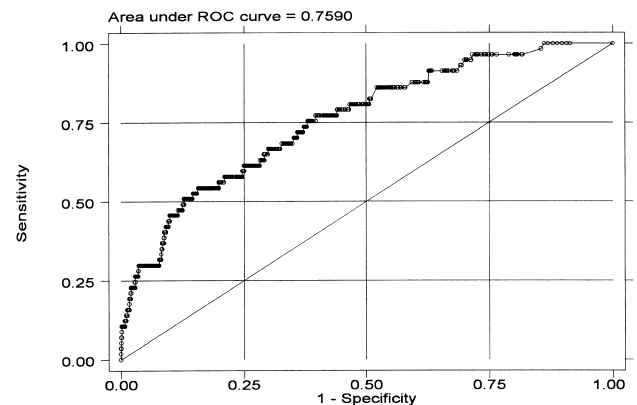


Fig. 2. ROC curve graphs for the validation dataset ($n = 1497$). The irregular form of the curve is due to the smaller sample size in the validation dataset.

Table 2
Risk factors, definitions and weights (score)

	Definition	Score
Patient-related factors		
Age	Per 5 years or part thereof over 60 years)	1
Sex	Female	1
Chronic pulmonary disease	Longterm use of bronchodilators or steroids for lung disease	1
Extracardiac arteriopathy	Any one or more of the following: claudication, carotid occlusion or >50% stenosis, previous or planned intervention on the abdominal aorta, limb arteries or carotids	2
Neurological dysfunction	Disease severely affecting ambulation or day-to-day functioning	2
Previous cardiac surgery	Requiring opening of the pericardium	3
Serum creatinine	>200 $\mu\text{mol/l}$ preoperatively	2
Active endocarditis	Patient still under antibiotic treatment for endocarditis at the time of surgery	3
Critical preoperative state	Any one or more of the following: ventricular tachycardia or fibrillation or aborted sudden death, preoperative cardiac massage, preoperative ventilation before arrival in the anaesthetic room, preoperative inotropic support, intraaortic balloon counterpulsation or preoperative acute renal failure (anuria or oliguria < 10 ml/h)	3
Cardiac-related factors		
Unstable angina	Rest angina requiring i.v. nitrates until arrival in the anaesthetic room	2
LV dysfunction	Moderate or LVEF 30–50%	1
	Poor or LVEF < 30	3
Recent myocardial infarct	(<90 days)	2
Pulmonary hypertension	Systolic PA pressure >60 mmHg	2
Operation-related factors		
Emergency	Carried out on referral before the beginning of the next working day	2
Other than isolated CABG	Major cardiac procedure other than or in addition to CABG	2
Surgery on thoracic aorta	For disorder of ascending, arch or descending aorta	3
Postinfarct septal rupture		4

Future debate in this field will focus on the quality of treatment and the measurement of this quality. In cardiac surgery, it has long been accepted that operative or hospital mortality is an indicator of quality of care. This is true to a large extent: death following heart surgery is often due to failure to achieve a satisfactory cardiac outcome, itself the

cause of major early morbidity as well as poor longterm results. Crude operative mortality fails as a measure of quality only when there are major variations in casemix. For operative mortality to remain a valid measure of quality of care, it must be related to the risk profile of the patients receiving surgery, hence the need for a reliable risk stratification model, already recognised by earlier workers in this field (see reference list of [1]).

There is another reason for the development and regular use of risk stratification in the assessment of cardiac surgical results. Doctors and hospitals operate in an increasingly open system where the availability of results and public accountability may influence decision-making. Without risk stratification, surgeons and hospitals treating high-risk patients will appear to have worse results than others. This may prejudice referral patterns, affect the allocation of resources and even discourage the treatment of high-risk patients. This is especially undesirable in cardiac surgery because it is precisely this group of patients which stands to gain most from surgical treatment, in spite of the increased risk [2]. Risk stratification helps eliminate the bias against high-risk patients [3].

An individual patient will either survive or die after cardiac surgery. Clearly, no scoring system will predict the specific outcome for every patient. Risk stratification, however, will inform patients and clinicians of the likely risk of death for a group of patients with a similar risk profile undergoing the proposed operation. This information is useful, and should form part of the basis on which the patient and surgeon decide whether to proceed.

The *EuroSCORE* database is large, up-to-date and unrivalled in completeness and accuracy. It is also derived from a cross-section of contemporary European cardiac surgery. It is therefore an appropriate database for the construction of a risk evaluation scoring system for use in Europe. The limitations of the study are those which are due to the study design in the areas of centre recruitment and data collection, and have already been addressed [1].

It can be argued that the transition from database to scoring system sacrifices some precision for the sake of simplicity. There are two extremes in the selection of a risk stratification system. Accuracy can be achieved by assessing a large number of risk factors for an individual patient and comparing the findings with the results of a large database such as *EuroSCORE*. Such a system should provide very accurate risk assessment for small subgroups of patients. This approach, however, would require the gathering of large amounts of patient data and complex statistical operations. It would be of limited use in the day-to-day world of clinical surgery, and impossible to implement without sophisticated information technology which is not yet available to all hospitals. On the other hand, very simple models relying on one or two risk factors (such as age and sex, for example) are also possible. This approach would have some use for the overall assessment of a hospital's performance, but is unlikely to be useful for risk assessment for an indi-

Table 3
Application of scoring system

EuroSCORE	Patients	Died	95% confidence limits for mortality	
			Observed	Expected
0–2 (low risk)	4529	36 (0.8%)	(0.56–1.10)	(1.27–1.29)
3–5 (medium risk)	5977	182 (3.0%)	(2.62–3.51)	(2.90–2.94)
6 plus (high risk)	4293	480 (11.2%)	(10.25–12.16)	(10.93–11.54)
Total	14799	698 (4.7%)	(4.37–5.06)	(4.72–4.95)

vidual patient and is likely to perpetuate a reluctance to operate on high-risk patients. A compromise must be reached so that the system recognises common risk factors, is able to provide some degree of risk prediction yet remains simple enough to use at the point of delivery of care [4,5]. *EuroSCORE* satisfies these requirements. The existence of the scoring system does not preclude full use of the database, when resources permit, for more precise analysis.

It is essential that the risk stratification system is objective and resistant to manipulation. This is achieved by the selection of real, measurable and easily available risk factors. In addition, it is important that as few risk factors as possible are determined by surgical decision-making. Most *EuroSCORE* risk factors are derived from the clinical status of the patient. Only four risk factors are related to the operation and these are factors that are difficult to influence through subtle variation in surgical decision-making.

EuroSCORE is sound in planning and derivation, easy to use and applicable either as a paper system or through information technology. However, the true test of such a system is in its widespread application in the field. We invite and welcome other workers to put it to the test in their hospitals, overall and in individual patient and procedural subgroups, in relation to operative mortality, to major morbidity and to the use of resources. Quality monitoring is now one of the requirements of good surgical practice: *EuroSCORE* is a tool by which this can be achieved.

References

- [1] Roques F, Nashef SAM, Michel P, Gauducheau E, de Vincentiis C, Baudet E, Cortina J, David M, Faichney A, Gabrielle F, Gams E, Harjula A, Jones MT, Pinna Pintor P, Salamon R, Thulin L. Risk factors and outcome in European cardiac surgery: analysis of the EuroSCORE multinational database of 19 030 patients. *Eur J Cardio-thorac Surg* 1999;15:816–823.
- [2] Mark DB. Implications of cost in treatment selection for patients with coronary disease. *Ann Thorac Surg* 1996;61:S12–S15.
- [3] Hannan FL, Siu AL, Kumar D, Racz M, Pryor DB, Chassin MR. Assessment of coronary artery bypass graft surgery performance in New York. Is there a bias against taking high risk patients? *Med Care* 1997;35:49–56.
- [4] Tu JV, Sykora K, Naylor CD. Assessing the outcomes of coronary artery bypass graft surgery: how many risk factors are enough? *J Am Coll Cardiol* 1997;30:1317–1323.
- [5] Jones RH, Hannan EL, Hammermeister KE, DeLong ER, O'Connor GT, Luepker RV, Parsonnet V, Pryor DB. Identification of preopera-

tive variables needed for risk adjustment of short-term mortality after coronary artery bypass graft surgery. *J Am Coll Cardiol* 1996;28:1478–1487.

Appendix A. Conference discussion

Dr T. Aberg (Umea, Sweden): When you compare this outcome with other risk stratification systems, what do you feel is the main difference?

Dr Nashef: We have done this in our own database, but then of course our risk stratification system is derived from that database and therefore it would be expected that it would perform well in that database, and this is one reason for my inviting other workers to take it on and compare it with other systems. We would very much like to see it tested.

Dr Aberg: I think this is one of the important papers of this meeting and you have heard the speaker make a plea for using the system more or less routinely. We participated in the study and we definitely are going to use it in conjunction with the usual Higgins and Parsonnet scores that we are currently collecting.

Dr Murday: Can I press you a little further to give your opinion as to what you see as being the advantages of your *EuroSCORE* system against, for example, Parsonnet, which a lot of us use already?

Dr Nashef: My opinion is likely to be biased because this is my work, but there are areas in which *EuroSCORE* is much more objective and Parsonnet is subjective. Parsonnet has been shown to overestimate risk and has had to be redone a number of times in order to correct for that. This will not be a problem with Euroscore. There are some risk factors that we simply have not found to be significant in spite of what Parsonnet has found. There are some risk factors that Parsonnet has excluded that all cardiac surgeons know are a serious contribution to risk, such as, for example, the presence of extracardiac vascular disease, and we have managed to show that impact. So we believe it is a significant advance. It does relate to European cardiac surgery because it is based on the European population.

Dr G. Rizzoli (Padua, Italy): I would like to know the composition of your database. How many patients were coronary bypass patients, how many were valvular patients? I would also like to know if you tested the results of the overall analysis on each subset of patients, the group with coronary artery bypass and the group with valvular disease?

Dr Nashef: Yes, we have. In the development of the score we looked at subgroups both as coronary patients and valve patients. As far as the composition is concerned and if I remember correctly, approximately 60% were coronary patients, about 30% valve patients and 10% other. We looked at developing separate scores for each population and we looked at developing an overall score, and we were successful in developing an overall score.

Dr F. Grover (Denver, CO, USA): I have had considerable experience working with our STS database in the United States and was curious about whether you thought about taking odds ratios for the various risk factors and rather than adding them up as a score, developing very simple software that you could have in a handheld computer which would allow you to estimate the risk of the patient and to separate into risk groups at 10% intervals or subsets.

Dr Nashef: So you mean to use the database itself as a backbone for predicting risk for individual patients?

Dr Grover: Yes.

Dr Nashef: Yes, that is certainly possible and we are looking at that in the future.

Dr Aberg: The publication and dissemination of this data, what is the time frame for you to do that? How do you plan to make these data known in more detail?

Dr Nashef: Well, we have expected it will be published in the *European Journal of Cardio-thoracic Surgery*.