

Evaluation of the EuroSCORE risk scoring model for patients undergoing coronary artery bypass graft surgery: a word of caution

A.H.M. van Straten, E.M.E.S.H. Tan, M.A.S. Hamad, E.J. Martens, A.A.J. van Zundert

Background. Risk-adjusted mortality rates are used to compare quality of care of different hospitals. We evaluated the EuroSCORE (European System for Cardiac Operative Risk Evaluation) in patients undergoing isolated coronary artery bypass grafting (CABG).

Patients and method. Data of all CABG patients from January 2004 until December 2008 were analysed. Receiver-operating characteristics (ROC) curves for the additive and logistic EuroSCOREs and the areas under the ROC curve were calculated. Predicted probability of hospital mortality was calculated using logistic regression analyses and compared with the EuroSCORE. Cumulative sum (CUSUM) analyses were performed for the EuroSCORE and the actual hospital mortality.

Results. 5249 patients underwent CABG of which 89 (1.7%) died. The mean additive EuroSCORE was 3.5 ± 2.5 (0-17) (median 3.0) and the mean logistic EuroSCORE was 4.0 ± 5.5 (0-73) (median 2.4). The area under

the ROC curve was 0.80 ± 0.02 (95% confidence interval (CI) 0.76 to 0.84) for the additive and 0.81 ± 0.02 (0.77 to 0.85) for the logistic EuroSCORE. The predicted probability (hazard ratio) was different from the additive and logistic EuroSCOREs. The hospital mortality was half of the EuroSCOREs, resulting in positive variable life-adjusted display curves.

Conclusions. Both the additive and logistic EuroSCOREs are overestimating the in-hospital mortality risk in low-risk CABG patients. The logistic EuroSCORE is more accurate in high-risk patients compared with the additive EuroSCORE. Until a more accurate risk scoring system is available, we suggest being careful when comparing the quality of care of different centres based on risk-adjusted mortality rates. (Neth Heart J 2010;18:355-9.)

Keywords: Coronary Artery Bypass; Risk Assessment; Outcome Assessment (Health Care); Quality-Adjusted Life Years; Health Status Indicators

A.H.M. van Straten

E.M.E.S.H. Tan

M.A.S. Hamad

Department of Cardiothoracic Surgery, Catharina Hospital, Eindhoven, the Netherlands

E.J. Martens

Department of Education and Research, Catharina Hospital, Eindhoven, and the Center of Research on Psychology in Somatic Diseases, Department of Medical Psychology, Tilburg University, the Netherlands

A.A.J. van Zundert

Department of Anaesthesiology, Catharina Hospital, Brabant Medical School, Eindhoven, the Netherlands and University Hospital Ghent, Ghent, Belgium

Correspondence to: P.A.H.M. van Straten
Department of Cardiothoracic Surgery, Catharina Hospital, PO Box 1350, 5602 ZA Eindhoven, the Netherlands
E-mail: cchbsn@cze.nl

In order to decide which therapeutic option (surgery, percutaneous intervention or medical therapy) is best for patients with coronary artery disease, a cost-benefit evaluation has to be performed for each therapeutic option. On the benefit side is the relief of symptoms and possible better long-term survival. On the cost side there is operative and postoperative morbidity and mortality. Several demographic data and preoperative, operative and postoperative variables (risk factors) may have impact on the results of the different therapies. Only when the impact of risk factors on morbidity and early and late mortality is known can a balanced decision about the best therapeutic option be made and a true informed consent from patients be obtained.

Table 1. Variables in the EuroSCORE.

Patient-related factors		Score	Beta
Age	Per 5 years or part thereof over 60 years	1	0.0666354
Sex	Female	1	0.3304052
Chronic pulmonary disease	Long-term use of bronchodilators or steroids for lung disease	1	0.4931341
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	0.6558917
Neurological dysfunction disease	Severely affecting ambulation or day-to-day functioning	2	0.841626
Previous cardiac surgery	Requiring opening of the pericardium	3	1.002625
Serum creatinine	>200 µmol/l preoperatively	2	0.6521653
Active endocarditis	Patient still under antibiotic treatment for endocarditis at the time of surgery	3	1.101265
Critical preoperative state	Any one or more of the following: VT or VF or aborted sudden death, cardiac massage, ventilation before arrival in the anaesthetic room, inotropic support, intra-aortic balloon counterpulsation or acute renal failure (anuria or oliguria <10 ml/hour)	3	0.9058132
Cardiac-related factors			
Unstable angina	Rest angina requiring iv nitrates	2	0.5677075
LV dysfunction	Moderate or LVEF 30-50%	1	0.4191643
	Poor or LVEF <30	3	1.094443
Recent MI	<90 days	2	0.5460218
Pulmonary hypertension	Systolic PA pressure >60 mmHg	2	0.7676924
Operation-related factors			
Emergency	Carried out on referral before the beginning of the next working day	2	0.7127953
Other than isolated CABG	Major cardiac procedure other than or in addition to CABG	2	0.5420364
Surgery on thoracic aorta	For disorder of ascending, arch or descending aorta	3	1.159787
Postinfarct septal rupture		4	1.462009

MI=myocardial infarction, VT=ventricular tachycardia, LVEF=left ventricular ejection fraction, PA=pulmonary artery, CABG=coronary artery bypass graft.

Importance of accurate risk scoring models for doctors and institutions

Nowadays, 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 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 important. 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 The hospital standardised mortality ratio (HSMR) as first described by Jarman et al.^{2,3} uses administrative data that are readily available in every

hospital such as year, age, sex, ICD-9 code, Charlson index for comorbidity, social-economic status of the patient, urgency, and length of stay in the hospital. The HSMR to measure quality of healthcare has been debated in the Dutch literature as it has many flaws and is considered inaccurate.⁴⁻¹³ 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 system,¹⁴ which was developed in the USA in the 1980s. Thereafter, many other risk scoring models were developed such as the Cleveland Clinic score,¹⁵ French score,¹⁶ Pons score¹⁷ and the Ontario Province score.¹⁸ Nowadays the most used risk stratification models are the EuroSCORE and the Society of Thoracic Surgeons (STS) score.

EuroSCORE

The European System for Cardiac Operative Risk Evaluation (EuroSCORE) was first described in 1999.^{19,20} Nearly 20,000 consecutive patients from 128 hospitals in eight European countries were studied. Information was collected on 97 risk factors in all patients. The outcome (survival or death) was related to the preoperative risk factors. In the newly developed Dutch national database the data from every Dutch cardiothoracic centre are collected including the EuroSCORE and the hospital mortality. The variables that are used to calculate the EuroSCORE are shown in table 1.

Two risk calculators are available: in the simple additive EuroSCORE the weights (score in table 1) are added to give an approximate percent predicted mortality. This simple additive EuroSCORE model is now well established and has been validated in many cardiac surgery patient populations across the world. It is very valuable in quality control in cardiac surgery and gives a useful estimate of risk in individual patients. However, particularly in very high-risk patients, the simple additive model may sometimes underestimate the risk when certain combinations of risk factors co-exist. The logistic version of EuroSCORE produces more accurate risk prediction for a particular high-risk patient using the 'Beta' from table 1 for its calculation. Its main disadvantage is that the risk has to be calculated in quite a complex way.

When the STS risk scoring model was compared with the EuroSCORE in patients undergoing coronary artery bypass graft surgery (CABG), Ad et al.²¹ found both models to be good predictors of operative mortality with a slight advantage for the STS model. However, Nilsson et al.²² found that the additive EuroSCORE had a better discriminatory power compared with the STS risk model.

The goal of this study was to evaluate the EuroSCORE as a risk stratification model for patients undergoing (CABG) surgery using the data we collected in our computerised database.

Patients and Method

From January 2004 we collected the EuroSCORE of all cardiac surgery patients in the Catharina Hospital, Eindhoven, the Netherlands. We analysed the data of all patients undergoing isolated CABG from January 2004 till December 2008. We measured the discriminatory power of the EuroSCORE for operative mortality. For this we calculated the receiver operating characteristics (ROC) curve and the area under the ROC curve. We calculated the cumulative summation of the EuroSCORE of all consecutive patients (CUSUM) and the cumulative hospital mortality as defined by mortality during the same hospital stay as the operation. From this we calculated the variable life-adjusted display

(VLAD) curve as described by Lovegrove et al. in 1997.²³ In constructing the VLAD curves, the EuroSCORE minus the mortality per patient are cumulatively summarised. Logistic regression analyses were used to calculate the predicted probability per EuroSCORE value, which reflects the actual hospital mortality.

Results

During the five-year study period, 5249 patients underwent CABG of which 89 (1.7%) died. The mean additive EuroSCORE was 3.5 ± 2.5 (0 to 17) and the mean logistic EuroSCORE was 4.0 ± 5.5 (0 to 73). The median additive EuroSCORE was 3.0 and the median logistic EuroSCORE was 2.4. The numbers of patients per additive EuroSCORE are shown in figure 1.

The ROC curves of both the additive and the logistic EuroSCORE are shown in figure 2. The area under the ROC curve was 0.80 ± 0.02 (95% confidence interval (CI) 0.76 to 0.84) for the additive EuroSCORE and 0.81 ± 0.02 (0.77 to 0.85) for the logistic EuroSCORE. The predicted probability (actual mortality) was different from the additive and the logistic EuroSCORE (figure 3). The cumulative additive EuroSCORE, logistic EuroSCORE and the hospital mortality are depicted in figure 4. The cumulative mortality is half of the predicted mortality according to the EuroSCORE. The VLAD curve was positive indicating that lives were spared (figure 5).

Discussion

In our patients, the EuroSCORE has a high discrim-

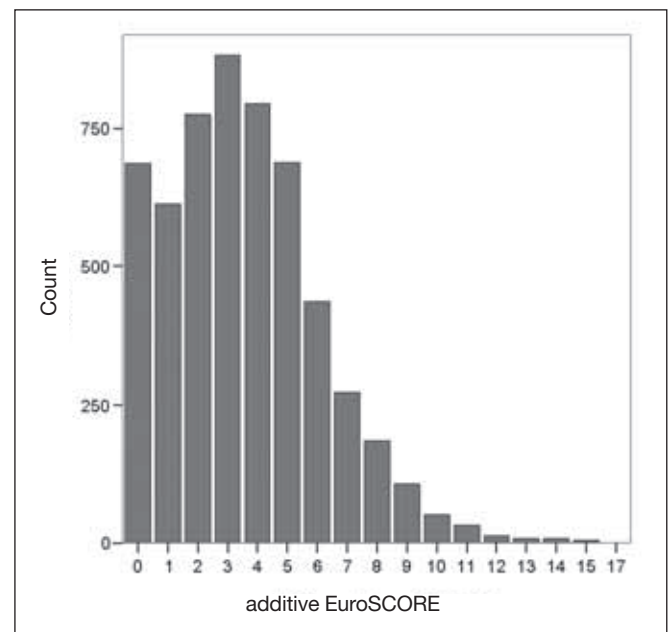


Figure 1. Distribution of additive EuroSCORE.

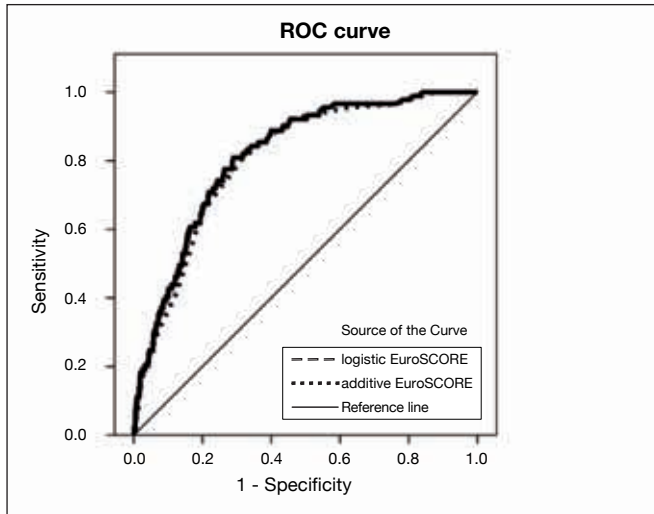


Figure 2. Receiver-operating characteristics (ROC) curve.

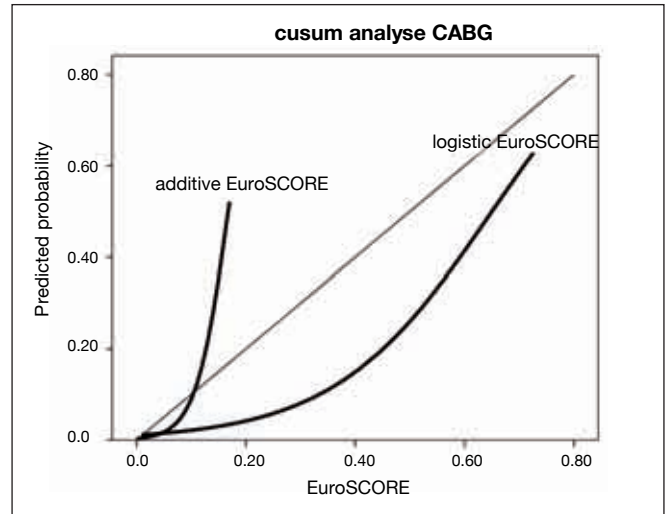


Figure 3. Predicted probability for hospital mortality and EuroSCORE.

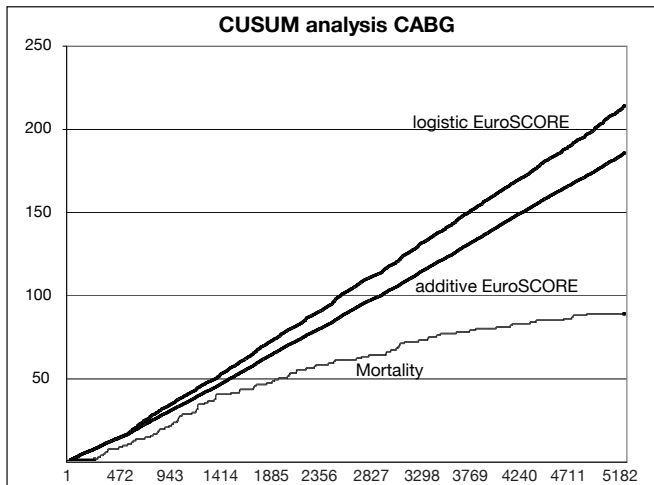


Figure 4. Cumulative additive, logistic EuroSCORE and hospital mortality.

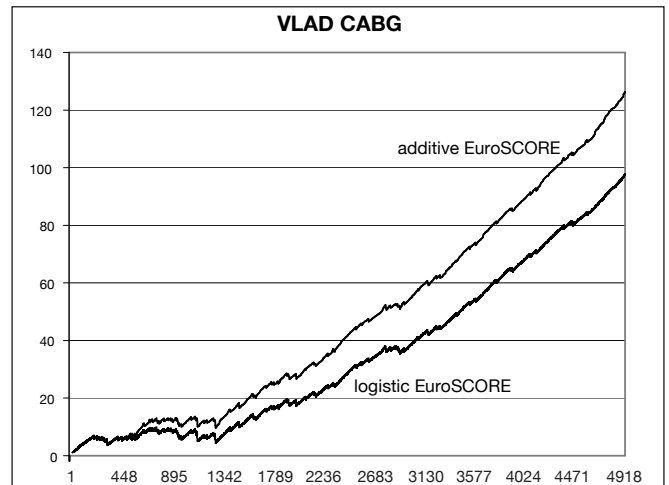


Figure 5. Variable life-adjusted display (VLAD) curves.

inatory power shown by an area under the ROC curve of 0.80 for the additive EuroSCORE and of 0.81 for the logistic EuroSCORE indicating that most patients who died had a high EuroSCORE. The area under the ROC curve is considered to be an indicator for the discriminatory power of the model. The discriminatory power is thought to be excellent if the area under the ROC curve is >0.80 , very good if >0.75 , and good if >0.70 .²³ Others found the similar results²² in a group of 4497 patients undergoing CABG. When they made a comparison with the STS score they found an area under the ROC curve of 0.85 (95% CI 0.80 to 0.88) for the EuroSCORE compared with 0.71 (95% CI 0.66 to 0.77) for the STS risk scoring system. They concluded that the EuroSCORE showed a larger discriminatory power compared with the STS risk

scoring system. Geisler et al.²⁴ reported an area under the ROC curve for the EuroSCORE of 0.78 in a group of 505 patients undergoing all kinds of heart surgery. When we plot the predicted probability calculated with logistic regression analyses in our CABG patients against the additive and logistic EuroSCORE (figure 3) it is clear that the logistic EuroSCORE is a better predictor of hospital mortality, especially in patients with a higher score. For lower logistic EuroSCOREs there was an over-estimation of the risk. The additive EuroSCORE underestimates the risk in high-risk patients. For a perfect risk scoring system the picture should show a diagonal line. One can conclude that the EuroSCORE model, though having a very good discriminatory power, is not accurate in predicting hospital mortality in CABG patients.

Recently, data on all cardiac surgery patients including the EuroSCORE, the EuroSCORE variables and the hospital mortality were collected in a Dutch national database. The CABG patients form an important and large group in this database. In order to compare the results of different centres using risk-adjusted mortality rates, we need reliable and accurate risk stratification models. If the calculated risk score is not accurate, comparison of risk-adjusted mortality may lead to the wrong conclusions. If, for example, the risk model underestimates the risk in high-risk patients and overestimates the risk in low-risk patients, as is the case in our CABG patients, the results of the comparison will always be in favour of the centres that are performing operations with the lowest risk scores. The VLAD curve will be more positive in centres with more low-risk patients. Improvement of the VLAD curve can be easily accomplished by denying operations to high-risk patients. Especially when the results of the VLAD curves are made public this issue will become more important. Re-evaluation of the variables used in the EuroSCORE and considering new ones in order to improve the accuracy of the scoring system is warranted. Only when a more accurate scoring system for CABG (and for all other cardiac operations) has been developed, as suggested by Nissinen et al.,²⁶ can publication of VLAD curves per centre be done more accurately without the risk that a hospital that is offering operations to the highest risk patients will be unfairly accused of 'bad quality'.

Conclusions

Both the additive EuroSCORE and the logistic EuroSCORE are overestimating hospital mortality risk in low-risk patients undergoing CABG. The logistic EuroSCORE is more accurate in high-risk patients compared with the additive EuroSCORE. Until a more accurate risk scoring system is available, we must be careful with comparing quality of care of different cardiac surgery centres based on risk-adjusted mortality rates. ■

References

- 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.
- Jarman B, Gault S, Alves B, Hider A, Dolan S, Cook A, et al. Explaining differences in English hospital death rates using routinely collected data. *BMJ*. 1999;318:1515-20.
- Jarman B, Bottle A, Aylin P, Browne M. Monitoring changes in hospital standardised mortality ratios. *BMJ*. 2005;330:329.
- van den Bosch WF, Roozendaal KJ, Silberbusch J. Schommellende sterftcijfers. *Medisch Contact*. 2009;31.
- Crommentuyn R. 'Levens redden wil iedereen'. Britten tonen dat verlagen sterftcijfer niet moeilijk is. *Medisch Contact*. 2009;26.
- Den Ouden AL, van der Wal G. Het gestandaardiseerde ziekenhuissterftcijfer (HSMR) bruikbaar voor het volgen van ziekenhuissterfte. *Ned Tijdschrift voor Geneeskunde*. 2008;152:1191-2.
- van den Bosch WF, Graafmans WC, Pieter D, Westert GP. Hartcentra en het effect van bijzondere medische verrichtingen op het gestandaardiseerde ziekenhuissterftcijfer. *Ned Tijdschr Geneesk*. 2008;152:1221-7.
- Den Ouden AL, Nugteren WAH, Haecck J, Van den Berg, MJM. Sterftcijfer niet afserven. *Medisch Contact*. 2008;63:561.
- Geelkerken RH, Mastboom WJB, Bertelink BP, Van der Palen J, Berg M, Kingma JH. Een onrijp instrument. Sterftcijfer niet geschikt als maat voor ziekenhuis kwaliteit. *Medisch Contact*. 2008;63:370-9.
- Kool RB, Van der Veen AA, Westert GP. Sterftemaat is valide instrument. *Medisch Contact*. 2007;62:2090-91.
- Van der Voort PHJ, De Jonge E. Sterfte als maat voor kwaliteit. *Medisch Contact*. 2007;62:1766-7.
- Pronk E. Mortaliteit als maat - Sterftcijfers geven kwaliteit ziekenhuiszorg weer. *Medisch Contact*. 2005;60: 876-9.
- Heijink R, Koolman X, Pieter D, Van der Veen A, Jarman B, Westert G. Measuring and explaining mortality in Dutch hospitals; the Hospital Standardized Mortality Rate between 2003 and 2005. *BMC Health Services Research*. 2008;8:73.
- 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.
- Higgins TL, Estafanous FG, Loop FD, Beck GJ, Blum JM, Parandhi L. Stratification of morbidity and mortality outcome by preoperative risk factors in coronary artery bypass patients. *J Am Med Assoc*. 1992;267:2344-8.
- Roques F, Gabrielle F, Michel P, de Vincentiis C, David M, Baudet E. Quality of care in adult heart surgery: proposal for a self-assessment approach based on a French multicenter study. *Eur J Cardiothorac Surg*. 1995;9:433-40.
- Pons JMV, Granados A, Espinas JA, Borrás JM, Martín I, Moreno V. Assessing open heart surgery mortality in Catalonia (Spain) through a predictive risk model. *Eur J Cardiothorac Surg*. 1997;11:415-23.
- Tu JV, Jaglal SB, Naylor CD. 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.
- Roques F, Nashef SAM, Michel P, Gauducheau E, de Vincentiis C, Baudet E, et al. Risk factors and outcome in European cardiac surgery: analysis of the EuroSCORE multinational database of 19030 patients. *Eur J Cardiothorac Surg*. 1999;15:816-23.
- Nashef SAM, Roques F, Michel P, Gauducheau E, Lemeshow S, Salamon R. European system for cardiac operative risk evaluation (EuroSCORE). *Eur J Cardiothorac Surg*. 1999;16:9-13.
- Ad N, Barnett SD, Speir AM. The performance of the EuroSCORE and the Society of Thoracic Surgeons mortality risk score: the gender factor. *Interact CardioVasc Thorac Surg*. 2007;7:192-5.
- Nilsson J, Algotsson L, Hoglund P, Luhrs C, Brandt J. Early mortality in coronary artery bypass surgery: the EuroSCORE versus the Society of Thoracic Surgeons risk algorithm. *Ann Thorac Surg*. 2004;77:1235-9.
- Lovegrove J, Valencia O, Treasure T, Sherlaw-Johnson C, Gallivan S. Monitoring the results of cardiac surgery by variable life-adjusted display. *Lancet*. 1997;350:1128-30.
- Swets JA. Measuring the accuracy of diagnostic systems. *Science*. 1988;240:1285-93.
- Geissler HJ, Holz P, Marohl S, Kuhn-Regnier F, Mehlhorn U, Sudkamp M, et al. Risk stratification in heart surgery: comparison of six score systems. *Eur J Cardiothorac Surg*. 2001;17:400-6.
- Nissinen J, Biancari F, Wistbacka JO, Lopenen P, Teittinen K, Tarkkiainen P, et al. Is it possible to improve the accuracy of EuroSCORE? *Eur J Cardiothorac Surg*. 2009;June 10 (Epub ahead of print).