

Mortality from isolated coronary bypass surgery: a comparison of the Society of Thoracic Surgeons and the EuroSCORE risk prediction algorithms

Irfan Qadir, Muhammad Musa Salick, Shazia Perveen and Hasanat Sharif*

Department of Cardiothoracic Surgery, Aga Khan University Hospital, Karachi, Pakistan

* Corresponding author. Department of Cardiothoracic Surgery, Aga Khan University Hospital, Stadium Road, Private Wing 2nd Floor, Karachi 74800, Pakistan. Tel: +92-21-34864708, +92-300-8238595; e-mail: hasanat.sharif@aku.edu (H. Sharif).

Received 18 July 2011; received in revised form 7 October 2011; accepted 18 October 2011

Abstract

We compared the performances of the additive European System for Cardiac Operative Risk Evaluation, EuroSCORE (AES) and logistic EuroSCORE (LES) with the Society of Thoracic Surgeons' risk prediction algorithm in terms of discrimination and calibration in predicting mortality in patients undergoing isolated coronary artery bypass grafting (CABG) at a single institution in Pakistan. Both models were applied to 380 patients, operated upon at the Aga Khan University Hospital from August 2009 to July 2010. The actual mortality was 2.89%. The mean AES of all patients was $4.36 \pm 3.58\%$, the mean LES was $5.96 \pm 9.18\%$ and the mean Society of Thoracic Surgeons' (STS) score was $2.30 \pm 4.16\%$. The Hosmer–Lemeshow goodness-of-fit test gave a *P*-value of 0.801 for AES, 0.699 for LES and 0.981 for STS. The area under the receiver operating characteristic curve was 0.866 for AES, 0.842 for LES and 0.899 for STS. STS outperformed AES and LES both in terms of calibration and discrimination. STS, however, underestimated mortality in the top 20% of patients having an STS score >2.88 , thus overall STS estimates were lower than actual mortality. We conclude that STS is a more accurate model for risk assessment as compared to additive and logistic EuroSCORE models in the Pakistani population.

Keywords: European System for Cardiac Operative Risk Evaluation • Society of Thoracic Surgeons • Coronary artery bypass grafting • Risk stratification • Pakistan

INTRODUCTION

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 long-term results. Crude operative mortality fails as a measure of quality when there are major variations in the case mix. It is widely accepted that monitoring of risk-adjusted mortality is one of the simplest methods for risk assessment, cost-benefit analysis and evaluation of hospital performance [1]. The growing interest in risk-adjusted analysis of outcome in cardiac surgery has led to the development and validation of several predictive models for post-operative mortality, morbidity and prolonged hospital stay in the two last decades [2].

The Society of Thoracic Surgeons' (STS) mortality risk score and the European System for Cardiac Operative Risk Evaluation (EuroSCORE) scoring system are the two most frequently used risk profile systems in cardiac surgery. The EuroSCORE scoring system for mortality comes in two versions: an AES and an LES [1]. It has been validated in individual European countries [3] and in Japan [4], Turkey [5] and North America [6]. The STS National Adult Cardiac Database, which is the largest of its kind, has been used to develop an algorithm to predict operative mortality [7].

The STS risk calculator offers the distinct advantage of predicting morbidity in terms of stroke, renal dysfunction, reoperation, prolonged ventilation, deep sternal wound infection and the length of hospital stay.

We recently tested the predictive performances of AES and LES in a single-centre retrospective study and found that both models over-predicted mortality at low (EuroSCORE 0–2) and medium risk levels (EuroSCORE 3–5). On the other hand, AES under-predicted and LES over-predicted mortality in high-risk patients (EuroSCORE >6) [8]. However, in the quest of a risk prediction algorithm better suited to Pakistani patients, we undertook this study to validate the STS risk algorithm in Pakistani patients and to compare its predictive performance with AES and LES in terms of discrimination and calibration on the same patient population.

PATIENTS AND METHODS

The Aga Khan University Hospital maintains a computerized database for all patients undergoing cardiac surgery since 2006. For this study, retrospective data were extracted on a subset of 380 patients who underwent isolated coronary artery bypass grafting (CABG) between August 2009 and July 2010. The EuroSCORE and STS risk scores were calculated with free online calculator

available at (<http://www.euroscore.org>) and (<http://209.220.160.181/STSWebRiskCalc261/de.aspx>), respectively.

Patient demographics were presented as percentages for discrete variables and mean (\pm SD) for continuous variables. Absolute mortality was determined for the overall patient population and trends in actual mortality were analysed across the entire risk spectrum. Performance of the models was also assessed by comparing the observed and expected mortality in quintiles of risk.

The performance of the AES, LES and STS risk algorithms were evaluated in terms of their discrimination and calibration. Discriminatory power was assessed using the area under the receiver operating characteristic (ROC) curve with 95% CI – an area of 0.5 indicates no predictive ability, whereas an area of 1.0 represents perfect discrimination [9]. Model calibration (the degree to which observed outcomes are similar to the predicted outcomes from the model across patients) of AES, LES and STS mortality and morbidity models was examined by comparing average observed and predicted values within each of 10 equal-

sized subgroups arranged in increasing order of patient risk. To evaluate model calibration, the Hosmer–Lemeshow (H–L) test for the lack of ‘goodness of fit’ was applied and graphically represented by a calibration plot [10]. The smooth curve in a calibration plot reflects the nonparametric relation between observed and predicted risk mortality. The straight dotted line through the origin of a calibration plot represents perfect calibration. H–L *P*-values above 0.05 indicate a well-calibrated model for the study population in question.

Statistical analyses were performed with SPSS software version 19.

RESULTS

The study population included 380 patients operated upon at the Aga Khan University Hospital. Mean age was 58.7 ± 9.4 years. The prevalence of various risk factors in study sample is shown in Table 1.

Predicted mortality was $4.36 \pm 3.58\%$ by AES, $5.96 \pm 9.18\%$ by LES and $2.30 \pm 4.16\%$ by STS. There were 11 deaths (2.89%) during the 30 day post-operative period. The specific predicted major morbidity rates included stroke (1.33% predicted vs. 0.3% actual), renal failure (3.84% predicted vs. 2.1% actual), reoperation (6.81% predicted vs. 2.4% actual), prolonged ventilation (13.26% predicted vs. 15.8% actual), and sternal infection (0.24% predicted vs. 0.3% actual).

Figure 1 shows the predicted mortality plotted against (a) actual mortality (b) AES (c) LES and (d) STS. Actual mortality remains low until EuroSCORE 9 on the additive model and rises up to a risk category of 17. In higher risk categories, actual mortality increases sharply and exceeds 50% at EuroSCORE 15. This figure also shows that AES over-predicts mortality till EuroSCORE of 10 whereas LES continues to over-predicts mortality across the entire range of patients. STS, on the other hand, over-predicts mortality till EuroSCORE of 4 and under-predicts thereafter.

Table 2 shows observed and predicted mortality in risk quintiles. Table 2 shows a good fit of STS estimates for the first four quintiles. From eighth decile onwards, corresponding to an STS score of >2.88 , mortality risk was always underestimated. On the other hand, AES continues to overestimate mortality till ninth decile, corresponding to AES of 10; thereafter, it under-estimates mortality. LES overestimates mortality across the entire range of patients in an incremental manner.

Figure 2 shows the ROC curves of the AES, LES and STS. The area under curve (AUC) was 0.866 for AES, 0.842 for LES and

Table 1: Prevalence of predictors in the reference and the study sample

	Study sample	EuroSCORE	STS
N	380	19 030	188 912
Age (mean+SD)	58.7	62.5	64.6
<60 years (%)	54.2	33.2	30.1
60–64 years (%)	16.1	17.8	14.1
65–69 years (%)	15.5	20.7	18.4
70–74 years (%)	9.2	17.9	18.3
>75 years (%)	5	9.6	19.1
Female (%)	15.8	27.8	30.9
Chronic pulmonary disease (%)	2.4	3.9	15.4
Extra cardiac arteriopathy (%)	0.5	11.3	19.0
Neurological dysfunction (%)	2.4	1.4	6.3
Previous cardiac surgery (%)	1.5	7.3	11.7
Serum creatinine >200 mmol/L (%)	3.2	1.8	2.1
Active endocarditis (%)	0	1.1	0.4
Critical preoperative condition(%)	38.7	4.1	9.0
Unstable angina (%)	50	8	21.7
LEVF 30–50 (%)	31.3	25.6	37.8
LEVF <30 (%)	16.8	5.8	5.2
Recent myocardial infarct (%)	0.6	9.7	20.9
Pulmonary hypertension (%)	–	2	5.7
Emergency (%)	12.4%	4.9	8.6
Other than isolated CABG (%)	–	36.4	18.8
Surgery on thoracic aorta (%)	–	2.4	0.9
Postinfarct septal rupture (%)	–	0.2	0.2

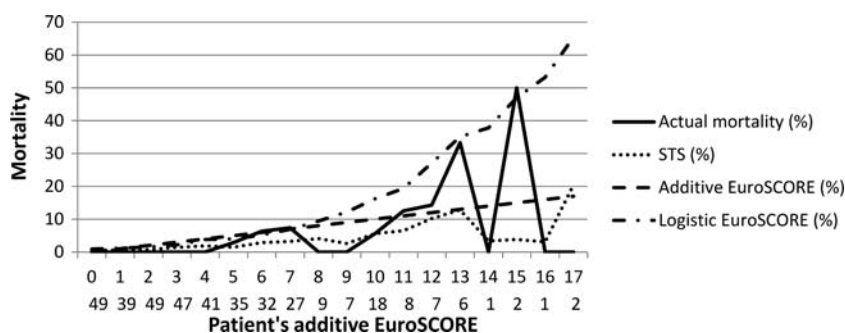


Figure 1: Predicted mortality (on the basis of additive EuroSCORE) plotted against (a) actual mortality (b) additive EuroSCORE (c) logistic EuroSCORE and (d) STS scores.

Table 2: Actual and predicted mortality by risk quintiles

	No. at risk	No. of observed death	Observed death (%)	Predicted death (%)
STS				
1st	85	0	0	0.32
2nd	82	0	0	0.66
3rd	69	0	0	1.41
4th	68	2	2.9	2.57
5th	76	9	11.8	10.61
AES				
1st	88	0	0	0.50
2nd	96	0	0	2.59
3rd	76	1	1.3	5.87
4th	59	4	6.8	10.95
5th	61	6	9.8	17.70
LES				
1st	92	0	0	1.04
2nd	60	0	0	2.60
3rd	76	0	0	3.49
4th	76	5	6.6	6.42
5th	76	6	7.9	25.96

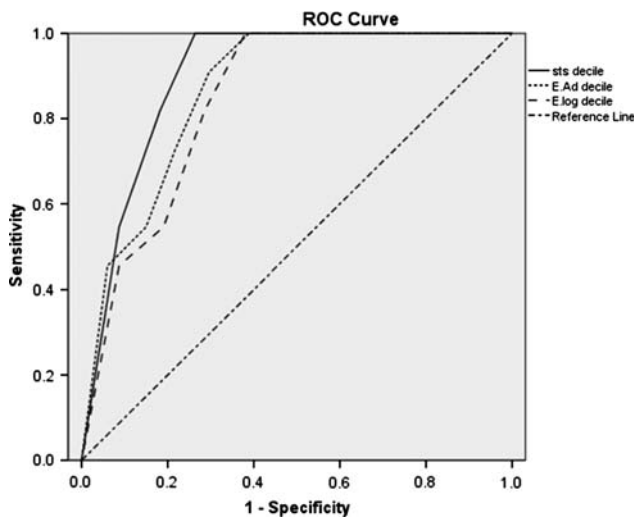


Figure 2: Receiver operating characteristic curve graphs for the AES, LES and STS scores.

0.899 for STS. Discriminatory power was significantly better for STS, which was demonstrated by a larger area under the ROC curve compared with AES ($P < 0.001$) and LES ($P < 0.001$).

Figure 3 demonstrates the calibration of models. The H-L P -value was 0.801 for AES, 0.699 for LES and 0.981 for STS 30 day mortality. The H-L test P -values for other STS outcome models are shown in Figure 3. STS was also better calibrated as noted by the close agreement between the actual and predicted event rate, the STS risk algorithm appeared to be relatively accurate across the entire range of patients.

DISCUSSION

Currently used risk-score systems have been developed for quite sometime and therefore require periodic re-calibration to reflect

improved surgical techniques and postoperative patient management advances which occurred in recent times. In addition, they are usually applied without validation to patient populations different from those from which they were derived. Differences in the prevalence of both measured and unmeasured variables (and performance characteristics) among the reference population and the testing sample, however, are generally considered as a serious hindrance to such a validation process [11]. Moreover, there are only a few studies comparing the performance STS and EuroSCORE [6, 12–14]. To our knowledge this is the first study which attempts to validate STS and compare it with other risk prediction algorithms in the Pakistani population.

The results in this study show that STS estimates are closer to observed rates compared to the EuroSCORE but that is to be expected; the STS prediction model is extremely comprehensive and includes 41 clinical variables versus 17 for the EuroSCORE. Secondly, better performance of STS can be explained by the periodic updates and revisions in the STS CABG risk models to reflect improved standards of cardiac care, the most recent of which was based upon 2002–2006 STS NCD data. This recalibration process included refinement, modification, consolidation or elimination of some data elements, so as to make the predicted mortality equal to the actual mortality derived from NCD [7].

However, STS significantly underestimates mortality in patients having a STS score >2.88 ; thus, overall STS estimates lower than actual mortality. According to the deciles distribution, the population of patients having a STS >2.88 can be defined as high-risk. Although high risk group (STS 2.88–37.2) constitutes a small percentage of the patients (20%) but it also express a significant proportion of the overall mortality (81.8%).

The difference in STS estimates and actual mortality are also explained by the difference in characteristics of Pakistani patients and American patients. The difference between the two populations is evident in our study wherein the mean age of patients undergoing surgery was 58.3 years as against 64.6 years in American population. About 54.3% of patients in our study were <60 years of age compared to only 33.2% patients with age <60 years in American population. Despite being younger, Pakistani patients have a higher prevalence of risk factors such as elevated serum creatinine, critical pre-operative condition, recent myocardial infarction, left ventricular dysfunction and unstable angina.

There are only a few studies comparing the performance of STS and EuroSCORE on the same patient population [6, 12–14]. The conclusion as to which model performs better remains controversial. Nilsson et al. [12] and Nashef et al. [6], after assessing the performance of STS and AES in Swedish population and STS national database, respectively, recommended the use of AES after demonstrating that the AES had better discriminatory power and predicted mortality remarkably similar to the observed mortality. In their study, Nashef et al. [6] stratified the STS data into quintiles of risk and calculated expected vs. observed mortality. In their commentary on this study, Mandel et al. [15] pointed out that the claim that EuroSCORE performs equally well on STS database is not based on statistical evidence. Our study has the advantage of subjecting the data to robust statistical analysis.

On the other hand, Pierri et al. [13] report that in Italian patients STS estimates (1.9%) were closer to actual mortality (1.9%) than AES (4.2%). Farrokhyar et al. [14] report that both models were equally good predictors of early mortality from off-pump and on-pump CABG in Canadian patients. The set of STS postoperative morbidity risk models also performed acceptably well on their data.

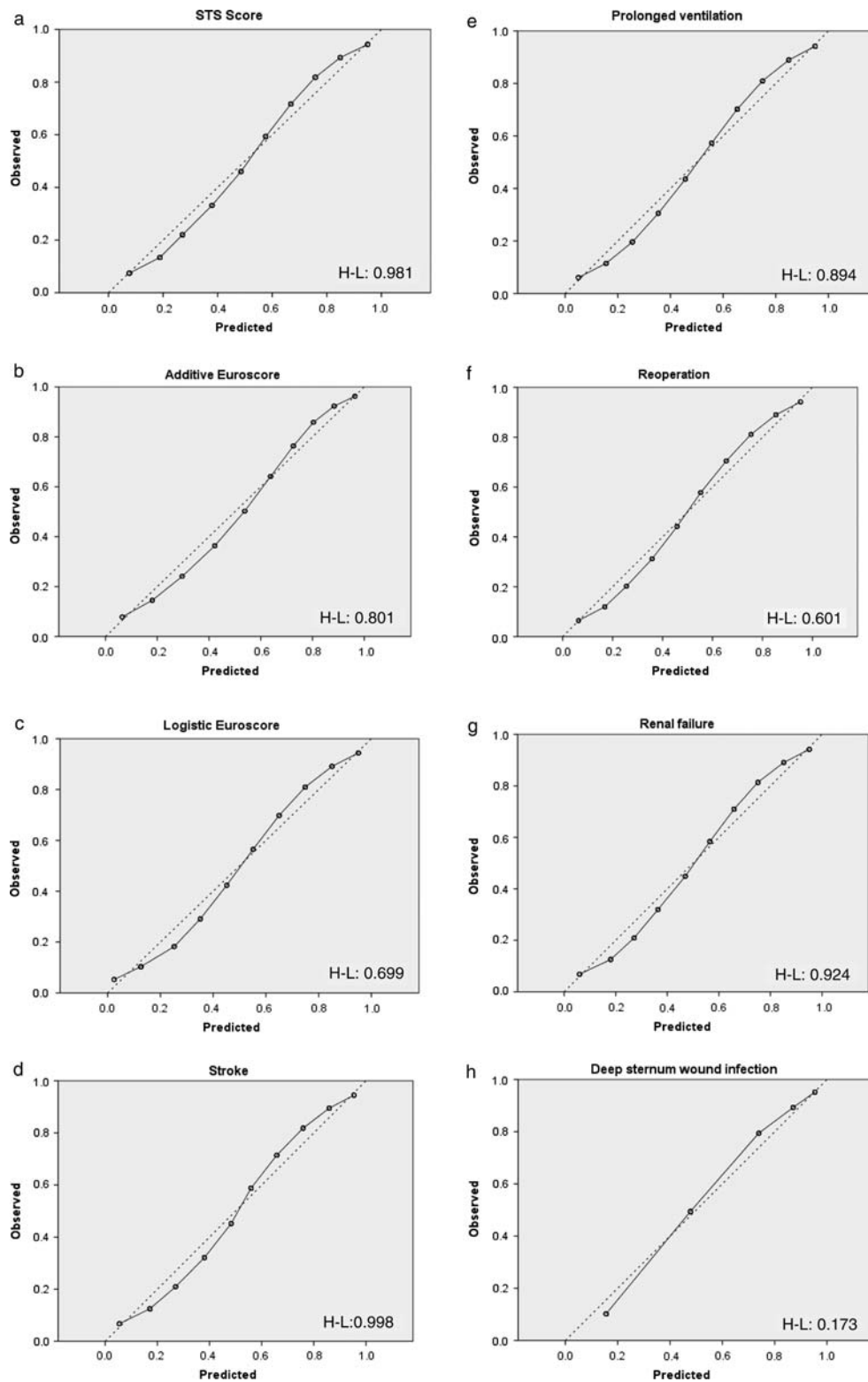


Figure 3: (a) Additive EuroSCORE calibration. (b) Logistic EuroSCORE calibration. (c) STS thirty-day operative mortality risk model calibration. (d) Stroke risk model calibration. (e) Reoperation risk model calibration. (f) Prolonged ventilation risk model calibration. (g) Renal failure risk model calibration. (h) Deep sternal wound infection risk model calibration.

LIMITATIONS

The small sample size and single-centre approach of this study limits the conclusions that can be drawn regarding the

rationale for which a risk stratification model should be applied in Pakistani patients. This should be borne in mind before extrapolating the findings to cardiac surgery in Pakistan as a whole.

CONCLUSION

This study concluded that the STS risk prediction algorithm is a better risk assessment tool compared to AES and LES in Pakistani patients.

ACKNOWLEDGEMENT

We would like to acknowledge Iqbal Azam's help with statistical analysis.

Conflict of interest: none declared.

REFERENCES

- [1] Nashef SA, 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.
- [2] Roques F, Nashef SA, 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;15:816-22.
- [3] Roques F, Nashef SA, Michel P, Pinna Pintor P, David M, Baudet E. Does EuroSCORE work in individual European countries? *Eur J Cardiothorac Surg* 2000;18:27-30.
- [4] Kawachi Y, Nakashima A, Toshima Y, Arinaga K, Kawano H. Evaluation of the quality of cardiovascular surgery care using risk stratification analysis according to the EuroSCORE additive model. *Circ J* 2002;66:145-8.
- [5] Karabulut H, Toraman F, Alhan C, Camur G, Evrenkaya S, Dagdelen S *et al.* EuroSCORE overestimates the cardiac operative risk. *Cardiovasc Surg* 2003;11:295-8.
- [6] Nashef SA, Roques F, Hammill BG, Peterson ED, Michel P, Grover FL *et al.* Validation of European System for Cardiac Operative Risk Evaluation (EuroSCORE) in North American cardiac surgery. *Eur J Cardiothorac Surg* 2002;22:101-5.
- [7] Shahian DM, O'Brien SM, Filardo G, Ferraris VA, Haan CK, Rich JB *et al.* The Society of Thoracic Surgeons 2008 cardiac surgery risk models: part 1—coronary artery bypass grafting surgery. *Ann Thorac Surg* 2009;88: S2-22.
- [8] Qadir I, Perveen S, Furnaz S, Shahabuddin S, Sharif H. Risk stratification analysis of operative mortality in isolated coronary artery bypass graft patients in Pakistan: comparison between additive and logistic EuroSCORE models. *Interact CardioVasc Thorac Surg* 13: 137-41.
- [9] Swets JA. Measuring the accuracy of diagnostic systems. *Science* 1988; 240:1285.
- [10] Hosmer DW, Lemeshow S. *Applied logistic regression*. New York: Wiley-Interscience, 2000.
- [11] Antunes PE, Eugenio L, Ferrao de Oliveira J, Antunes MJ. Mortality risk prediction in coronary surgery: a locally developed model outperforms external risk models. *Interact CardioVasc Thorac Surg*. 2007; 6:437-41.
- [12] Nilsson J, Algotsson L, Hoglund P, Luhrs C, Brandt J. Early mortality in coronary bypass surgery: the EuroSCORE versus The Society of Thoracic Surgeons risk algorithm. *Ann Thorac Surg* 2004; 77:1235-9.
- [13] Pierri MD, Borioni M, Iacobone G, Piccoli GP, Di Eusanio M, Bianchini F *et al.* [Mortality risk estimation in cardiac surgery]. *Ital Heart J Suppl* 2004;5:137-41.
- [14] Farrokhvar F, Wang X, Kent R, Lamy A. Early mortality from off-pump and on-pump coronary bypass surgery in Canada: a comparison of the STS and the EuroSCORE risk prediction algorithms. *Can J Cardiol* 2007; 23:879-83.
- [15] Mandel M, Simchen ES, Zitser-Gurevich Y. Does the EuroSCORE perform well on the STS population? *Eur J Cardiothorac Surg*. 2003;24: 336-7.