Interactive CardioVascular and Thoracic Surgery 22 (2016) 141–148
doi:10.1093/icvts/ivv304 Advance Access publication 6 November 2015

Cite this article as: Madeira S, Rodrigues R, Tralhão A, Santos M, Almeida C, Marques M *et al*. Assessment of perioperative mortality risk in patients with infective endocarditis undergoing cardiac surgery: performance of the EuroSCORE I and II logistic models. Interact CardioVasc Thorac Surg 2016;22:141–8.

Assessment of perioperative mortality risk in patients with infective endocarditis undergoing cardiac surgery: performance of the EuroSCORE I and II logistic models

Sérgio Madeira[a,*], Ricardo Rodrigues[b], António Tralhão[a], Miguel Santos[a], Carla Almeida[a], Marta Marques[c], Jorge Ferreira[a], Luís Raposo[a], José Neves[c] and Miguel Mendes[a]

[a] Department of Cardiology, Hospital de Santa Cruz, Lisbon, Portugal
[b] Department of Cardiology, Hospital Central do Funchal, Funchal, Madeira, Portugal
[c] Department of Cardiothoracic Surgery, Hospital de Santa Cruz, Lisbon, Portugal

* Corresponding author. Avenida Prof. Reinaldo dos Santos, 2790-134 Carnaxide, Portugal. Tel: +351-912-432444; e-mail: serg.lou.madeira@gmail.com (S. Madeira).

Received 27 June 2015; received in revised form 21 September 2015; accepted 28 September 2015

Abstract

OBJECTIVES: The European System for Cardiac Operative Risk Evaluation (EuroSCORE) has been established as a tool for assisting decision-making in surgical patients and as a benchmark for quality assessment. Infective endocarditis often requires surgical treatment and is associated with high mortality. This study was undertaken to (i) validate both versions of the EuroSCORE, the older logistic EuroSCORE I and the recently developed EuroSCORE II and to compare their performances; (ii) identify predictors other than those included in the EuroSCORE models that might further improve their performance.

METHODS: We retrospectively studied 128 patients from a single-centre registry who underwent heart surgery for active infective endocarditis between January 2007 and November 2014. Binary logistic regression was used to find independent predictors of mortality and to create a new prediction model. Discrimination and calibration of models were assessed by receiver-operating characteristic curve analysis, calibration curves and the Hosmer–Lemeshow test.

RESULTS: The observed perioperative mortality was 16.4% ($n = 21$). The median EuroSCORE I and EuroSCORE II were 13.9% interquartile range (IQ) (7.0–35.0) and 6.6% IQ (3.5–18.2), respectively. Discriminative power was numerically higher for EuroSCORE II {area under the curve (AUC) of 0.83 [95% confidence interval (CI), 0.75–0.91]} than for EuroSCORE I [0.75 (95% CI, 0.66–0.85), $P = 0.09$]. The Hosmer–Lemeshow test showed good calibration for EuroSCORE II ($P = 0.08$) but not for EuroSCORE I ($P = 0.04$). EuroSCORE I tended to over-predict and EuroSCORE II to under-predict mortality. Among the variables known to be associated with greater infective endocarditis severity, only prosthetic valve infective endocarditis remained an independent predictor of mortality [odds ratio (OR) 6.6; 95% CI, 1.1–39.5; $P = 0.04$]. The new model including the EuroSCORE II variables and variables known to be associated with greater infective endocarditis severity showed an AUC of 0.87 (95% CI, 0.79–0.94) and differed significantly from EuroSCORE I ($P = 0.03$) but not from EuroSCORE II ($P = 0.4$).

CONCLUSIONS: Both EuroSCORE I and II satisfactorily stratify risk in active infective endocarditis; however, EuroSCORE II performed better in the overall comparison. Specific endocarditis features will increase model complexity without an unequivocal improvement in predictive ability.

Keywords: Infective endocarditis • Heart surgery • Risk stratification

INTRODUCTION

Infective endocarditis (IE) is a dynamic disease constantly changing over time from an epidemiological, microbiological, diagnostic and therapeutic standpoint. However, the prognosis remains poor, especially when surgery is required [1]. Short-term mortality rates have been reported to range from 9% [2] in elective patients to 25–36% for those undergoing urgent surgery [2, 3]. Thus, risk stratification is of paramount importance not only for decision-making, but also for patient counselling (ensuring genuine informed consent) and comparative assessment of quality of care. Risk stratification and prediction models derived from large populations have gained popularity and are used worldwide for this purpose [4].

The European System for Cardiac Operative Risk Evaluation (EuroSCORE) was developed to predict procedure-related mortality rate within the first 30 days or during the initial hospitalization in adults undergoing heart surgery. The EuroSCORE-I (ES-I) was

initially created in 1999 [5] as an additive model and since 2003 it has been also available in a logistic version [6]. Only 202 (1.1%) patients of the 19 030 included in the derivation cohort had active IE [5]. Since then, its predictive performance (mostly concerning calibration) has been decreasing due to the changing epidemiology of cardiac surgery case mix and the improvement of surgical techniques and of postoperative care [4, 7, 8]. To overcome these limitations, an updated version, the EuroSCORE-II (ES-II), was modelled from a contemporary surgical cohort of 22381 patients, including 497 (2.2%) with active IE [9].

Nowadays, both ES-I and ES-II are both well accepted and used routinely in clinical practice in many countries [4, 7, 9]. However, given the singular characteristics of patients with IE that requires surgical management and the fact that these patients were underrepresented in the derivation cohorts of the prediction models, firm data are still lacking on their external reliability, predictive and discriminative ability in this setting. The present study sought to (i) externally validate both versions of the ES and to compare their performances; (ii) identify predictors other than those included in the ES models that might further improve their performance.

METHODS

Study design

The study population consisted of all the 128 patients who underwent cardiac surgery under extracorporeal circulation for active IE (according to modified Duke criteria) [10] between January 2007 and November 2014 in a single tertiary centre. Demographic, clinical, laboratorial, procedure-related information and perioperative vital status were initially retrieved from all patients from a dedicated institutional database. When unavailable, data were completed with information from clinical files, telephone contact or by consulting the National Social Security database, for vital status. ES-I and the new ES-II were calculated in accordance with published guidelines using a dedicated online calculator. As both ES models were primarily validated for perioperative mortality prediction, the only outcome assessed was mortality at 30 days or during the index hospitalization. All analyses performed were retrospective.

Statistical analysis

Data analysis. Normality was tested with the Kolmogorov–Smirnov test and/or Q-Q Plot visual assessment. Continuous variables with normal distribution were expressed as means and standard deviation and those without normal distribution as median and interquartile range. Discrete variables were expressed as frequencies and percentages. Statistical comparison of baseline characteristics was performed using Fisher's exact test for categorical variables, and Student's *t*-test or the Mann–Whitney *U*-test for continuous variables, when appropriate. Logistic regression was used to identify independent predictors of mortality among variables known to be associated with greater severity of IE that were not included in the ES models.

Model performance. The relative performance of the different ES models was compared using three different statistical methods: (i) discriminative power, using receiver-operating characteristic (ROC) curve analysis; (ii) calibration, using calibration curves of predicted versus observed mortality and the Hosmer–Lemeshow goodness-of-fit test and (iii) accuracy, using the Brier score of model residuals.

Discrimination indicates the extent to which the model distinguishes between patients who will die or survive in the perioperative period. It was evaluated by constructing ROC curves for each model and calculating the area under the curve (AUC) with 95% confidence intervals (CIs). The comparison between curves was assessed with the method described by Delong *et al.* [11]. Calibration refers to the agreement between observed outcomes and predictions and was evaluated by using calibration curves and the Hosmer–Lemeshow goodness-of-fit test. Calibration curves were constructed by plotting predictions in the *X*-axis and the observed outcome in the *Y*-axis (by each decile of the score-derived predictions). Subsequently, a linear regression was applied to the plot and a trend line was inferred. The resulting plots allow for a visual comparison between the predicted and the observed probability of the outcome and are characterized by an intercept, which indicates the extent to which predictions are systematically low or high, and a calibration slope that should be 0. The perfectly calibrated predictions stay on the 45° line, whereas a curve below or above the diagonal reflects over- and under-prediction, respectively. Furthermore, calibration was tested with the Hosmer–Lemeshow goodness-of-fit test, which compares observed with predicted values by decile of predicted probability.

The accuracy of the models was also tested calculating the Brier score [the quadratic difference between predicted probability and observed binary outcome (0 for no event and 1 for event) for each patient]. It is an overall performance measure that ranges between 0 and 1 with lower values indicating better performance.

All tests were two-sided and differences were considered statistically significant at a *P*-value of 0.05. Statistical analysis was performed with the SPSS 20.0 software (SPSS, Inc., Chicago, IL, USA) and MedCalc version 9.3.8.0 (MedCalc ©, Acacialaan Ostend, Belgium).

RESULTS

Population and surgery features

During the study time frame, 128 patients were submitted to heart surgery due to active IE. Ninety-six patients (75%) were male and the median age was 60 interquartile range (IQ) (47–70) years. Almost 80% of the patients had some kind of renal impairment including 14 patients (11%) on dialysis before surgery; 22% ($n = 28$) had left ventricular systolic dysfunction; ∼60% ($n = 62$) were in NYHA functional class III or IV at the time of surgery. Demographic and clinical baseline characteristics are summarized in Tables 1 and 2, for variables included and not included in the EuroSCOREs, respectively.

Surgery was considered urgent in 92% ($n = 117$) of the cases, emergent in 6% ($n = 8$) and salvage in 2% ($n = 3$). The main reasons for surgery were refractory heart failure due to valvular dysfunction and persistent infection. The description of the interventions is given in Table 3.

Specific endocarditis-related features (type of endocarditis, affected valves and causal agent) are described in Table 4.

Mortality

Observed mortality was 16% ($n = 21$), the median (IQ) time from surgery until the fatal event was 4 (1.75–30) days and 78% of deaths occurred within 30 days.

Patients who died within 30 days after surgery or during the index hospital admission were significantly older [65 IQ (57–74) vs

Table 1: Baseline characteristics included in the ES models

	Total (*n* = 128)	Alive (*n* = 107)	Perioperative mortality (*n* = 21)	*P*-value*
Patient-related factors				
Age (years)	60 IQ (47–70)	58 IQ (45–69)	65 IQ (57–74)	0.02
Male sex	96 (75%)	82 (77%)	14 (67%)	0.41
Extracardiac arteriopathy	10 (7.8%)	7 (6.5%)	3 (14.3%)	0.21
Poor mobility	12 (9.4%)	11 (10.3%)	1 (4.8%)	0.69
Previous cardiac surgery	30 (23.4%)	17 (15.9%)	13 (62%)	<0.001
Renal function				
Creatinine clearance >85 ml/min	26 (20%)	25 (23.4%)	1 (4.8%)	0.07
Creatinine clearance 50–85 ml/min	42 (33%)	37 (35%)	5 (32.8%)	0.45
Creatinine clearance <50 ml/min	46 (36%)	34 (32%)	12 (57%)	0.04
Serum creatinine >200 mmol/l	37 (28.9%)	29 (27%)	8 (38%)	0.3
Dialysis	14 (11%)	11 (10%)	3 (14.3%)	0.7
Chronic lung disease	10 (7.8%)	9 (8.4%)	1 (4.8%)	0.35
Critical preoperative state	14 (11%)	9 (8.4%)	5 (24%)	0.05
Diabetes	21 (16.4%)	15 (14%)	6 (28.6%)	0.11
Diabetes on insulin	6 (4.7%)	4 (3.7%)	2 (9.5%)	0.25
Cerebrovascular disease	4 (3.1%)	3 (2.8%)	1 (4.8%)	1
LV ejection fraction				
Good (LVEF >50%)	100 (78%)	86 (80.4%)	14 (66.7%)	0.25
Moderate (LVEF 31–50%)	22 (17.2%)	17 (15.9%)	5 (23.8%)	0.36
Poor (LVEF 21–30%)	6 (4.7%)	4 (3.7%)	2 (9.5%)	0.26
Very poor (LVEF <20%)	0	0	0	NA
CCS class 4 angina	2 (1.6%)	1 (0.9%)	1 (4.8%)	0.3
Recent MI <90 days	3 (2.3%)	1 (0.9%)	2 (9.5%)	0.07
NYHA				
NYHA IV	38 (29.7%)	32 (30%)	6 (28.6%)	1.0
NYHA III	39 (30.5%)	30 (28%)	6 (28.6%)	0.2
NYHA II	26 (20.3%)	23 (21.5%)	3 (14.3%)	0.56
NYHA I	25 (19.5%)	22 (20.6%)	3 (14.3%)	0.76
Pulmonary hypertension				
Moderate (PASP 31–55 mmHg)	9 (7%)	7 (6.5%)	2 (9.5%)	0.64
Severe (PASP >55 mmHg)	15 (11.7%)	13 (12.1%)	2 (16.7%)	1.0
EuroSCORE I (%)	24.4 ± 22.8	21.3 ± 20.6	40.1 ± 27.1	<0.001
EuroSCORE II (%)	11.9 ± 12.9	9.3 ± 9.7	25.2 ± 18.6	<0.001

**P*-values are for the comparison between patients who died during the index admission or at 30 days versus those who survived.
ES: EuroSCORE; EuroSCORE: European System for Cardiac Operative Risk Evaluation; IQ: interquartile range; CCS: Canadian Cardiovascular Society grading system for stable angina; LVEF: left ventricle ejection fraction; MI: myocardial infarction; NYHA: New York Heart Association functional class; PASP: pulmonary artery systolic pressure.

Table 2: Baseline and laboratory characteristics not included in the ES models

	Total (*n* = 128)	Alive (*n* = 107)	Perioperative mortality (*n* = 21)	*P*-value*
Patient-related factors				
Body mass index (kg/m^2)	25 ± 5.2	25 ± 5.2	26 ± 5.3	0.28
Underlying heart disease	56 (43.8%)	42 (39.3%)	14 (66.7%)	0.03
Laboratory				
Haemoglobin (g/dl)	9.7 (8.8–10.6)	9.7 (8.9–10.6)	9.6 (8.5–10.2)	0.54
White blood cell	8450 (6675–13 425)	8200 (6600–12 600)	12 500 (7050–14 650)	0.11
C-reactive protein (mg/dl)	5.3 (2.2–12.6)	5.2 (2–10)	7 (3.4–16.9)	0.11
Platelets	238 (158–308)	234 (157–302)	262 (159–345)	0.26
Haemoglobin <10[a] g/dl	75 (58.6%)	60 (56%)	15 (71.4%)	0.23
White blood cell >12 400[a]	40 (31.3%)	29 (27%)	11 (52.4%)	0.04
C-reactive protein >3[a] mg/dl	89 (69.5%)	71 (66.4%)	18 (85.7%)	0.12
Platelets <297[a]	42 (33%)	32 (30%)	10 (47.6%)	0.13

[a]Best discriminative value for operative mortality by ROC curve analysis.
**P*-values are for the comparison between patients who died during the index admission or at 30 days versus those who survived.

58 IQ (45–69) years old; $P = 0.02$], presented more frequently with perivalvular complications (48 vs 24%; $P = 0.04$), critical perioperative state (24 vs 8%; $P = 0.05$), creatinine clearance <85 ml/min (95 vs 75%; $P = 0.04$) and with higher levels of white blood cell count (>12.4 × 10^9/l, cut-off determined by c-statistics), had a higher prevalence of underlying heart disease (67 vs 39%;

Table 3: Surgical features

	Total (*n* = 128)	Alive (*n* = 107)	Perioperative mortality (*n* = 21)	*P*-value*
Reason for surgery				
Refractory heart failure due to valvular dysfunction	95 (74.2%)	81 (75.7%)	14 (66.7%)	0.42
Persistent infection	64 (50%)	52 (48.6%)	12 (57.1%)	0.63
Recurrent embolism	21 (16.4%)	18 (16.8%)	3 (14.3%)	1
Perivalvular complications	56 (28.1%)	26 (24.3%)	10 (47.6%)	0.04
Urgency				
Elective	0	0	0	NA
Urgent	117 (91.4%)	101 (94.4%)	16 (76.7%)	0.02
Emergent	8 (6.3%)	6 (5.6%)	2 (9.5%)	0.62
Salvage	3 (2.3%)	0	3 (14.3%)	0.004
Weight of the intervention				
Isolated CABG	0	0	0	NA
Single non-CABG	104 (81.3%)	86 (80.4%)	18 (85.7%)	0.76
Two procedures	23 (18%)	21 (19.6%)	2 (9.5%)	0.36
Three procedures	1 (0.8%)	0	1 (4.8%)	0.16
Surgery on thoracic aorta	4 (3.1%)	4 (3.7%)	0	1.0
Duration of hospital admission in our institution (days)	22 (9–44)	22 (9–63)	11 (2–43)	0.07

*P-values are for the comparison between patients who died during the index admission or at 30 days versus those who survived.
CABG: coronary artery by-pass grafting; NA: not applicable.

Table 4: Endocarditis-related features

	Total (*n* = 128)	Alive (*n* = 107)	Perioperative mortality (*n* = 21)	*P*-value*
Type of endocarditis				
Native valve	94 (73.4%)	81 (75.7%)	7 (33.3%)	<0.001
Prosthetic	29 (22.7%)	16 (15%)	13 (62%)	<0.001
Intracardiac device	5 (3.9%)	5 (4.6%)	0 (0%)	1
Involved structures				
Aortic valve	64 (50%)	57 (53.3%)	7 (33.3%)	0.15
Mitral valve	66 (51.6%)	52 (48.6%)	14 (66.7%)	0.15
Tricuspid valve	10 (7.8%)	8 (7.5%)	2 (9.5%)	0.67
Multivalvular	20 (15.6%)	16 (15%)	4 (19%)	0.74
Causal agents				
Streptococcus species	24 (19%)	23 (21.7%)	1 (4.8%)	0.12
Staphylococcus species	24 (18.8%)	18 (16.8%)	6 (28.6%)	0.23
Enterococcus	15 (11.7%)	14 (13.1%)	1 (4.8%)	0.46
Gram-negative bacteria	15 (11.7%)	12 (11.2%)	3 (14.3%)	0.71
Fungus	5 (3.9%)	4 (3.7%)	1 (4.8%)	1
Not identified	42 (32.8%)	32 (29.9%)	10 (47.6%)	0.13

*P-values are for the comparison between patients who died during the index admission or at 30 days versus those who survived.

$P = 0.03$), prior heart surgery (62 vs 16%; $P < 0.001$), prosthetic (versus native valve) endocarditis (62 vs 15.0%; $P < 0.001$) and underwent salvage surgery more frequently (14 vs 0.0%; $P = 0.004$). Of these, only the type of endocarditis (prosthetic versus native), perivalvular complications, underlying heart disease and white blood cell count are not included in the ES scoring systems. These variables were analysed in a multivariable regression including the ES-I and ES-II models, separately. In the analysis performed with ES-II, only this model and prosthetic valve IE proved to be independent predictors of mortality [odds ratio (OR) 1.06; 95% CI, 1.01–1.105; $P = 0.01$ and OR 6.6; 95% CI, 1.1–39.5; $P = 0.04$, respectively]. In the model including ES-I, only prosthetic valve IE preserved predictive ability (OR 7.6; 95% CI, 1.3–43; $P = 0.02$).

Performance of the EuroSCORE logistic models

Discriminative power. The mean values of the ES-I and II were 24.4 ± 22.8 and 11.9 ± 12.9, respectively (Table 1). The area under the ROC curve for the ES-I was 0.75 (95% CI, 0.66–0.85) and 0.83 (95% CI, 0.75–0.91) for the ES-II. Despite being numerically superior for the ES-II, the difference was not statistically significant (DeLong test, $P = 0.094$) (Table 5). The 'new' model including the ES-II variables and variables known to be associated with greater IE severity (modified ES-II) showed an AUC of 0.87 (95% CI, 0.79–0.94), and differed significantly from ES-I (DeLong test, $P = 0.03$) but not from ES-II (DeLong test, $P = 0.4$).

The ROC curves for all the models with respect to the study end-point are depicted in Fig. 1.

Table 5: Predictive performance-related statistics

	EuroSCORE I	EuroSCORE II	Modified EuroSCORE II
Overall performance			
Brier score	0.13	0.11	0.10
Discrimination			
AUC (95% CI)	0.75 (0.66–0.85)	0.83 (0.75–0.91)	0.87 (0.79–0.94)
Calibration			
Slope	0.49	1.1	0.8
Intercept	4.2	2.9	1.9
Hosmer–Lemeshow test, *P*-value	0.04	0.08	0.287
R^2 Nagelkerke	0.136	0.263	0.318
χ^2	15.9	13.9	9.7

AUC: area under the curve; CI: confidence interval.

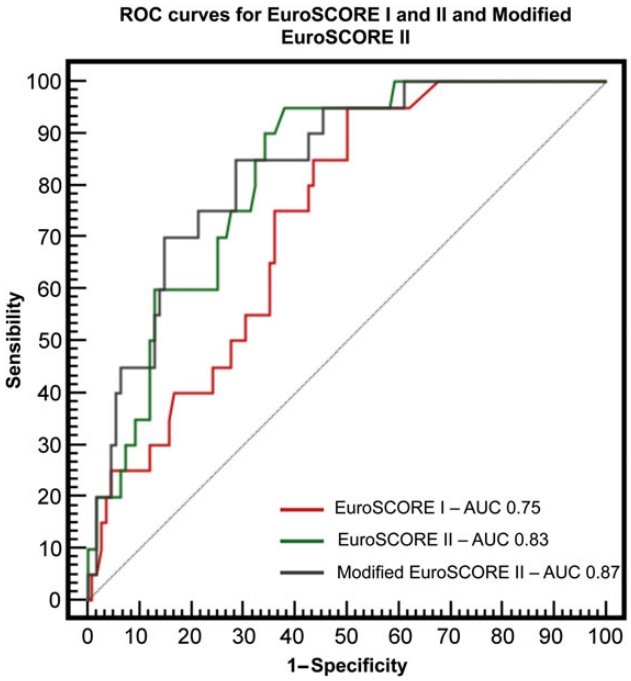

Figure 1: ROC curves for EuroSCORE I, EuroSCORE II and modified EuroSCORE II. EuroSCORE: European System for Cardiac Operative Risk Evaluation.

Calibration and accuracy. The calibration curves of ES I, ES-II and modified ES-II are shown in Fig. 2 The pattern of calibration was different between the scores: the ES-I showed a progressive trend towards over-prediction; on the other hand, ES-II and modified ES-II tended to under-predict mortality.

The calibration curve slope and intercept for each model is summarized in Table 5. ES-II and modified ES-II had non-significant *P*-values ($P = 0.08$ for ES-II and $P = 0.28$ for modified ES-II) for the Hosmer–Lemeshow test indicating that they would provide accurate probabilities whereas ES-I showed poor calibration ($P = 0.04$) (Table 5). The overall performance ascertained by the Brier score was good, with values near to 0 (Table 5).

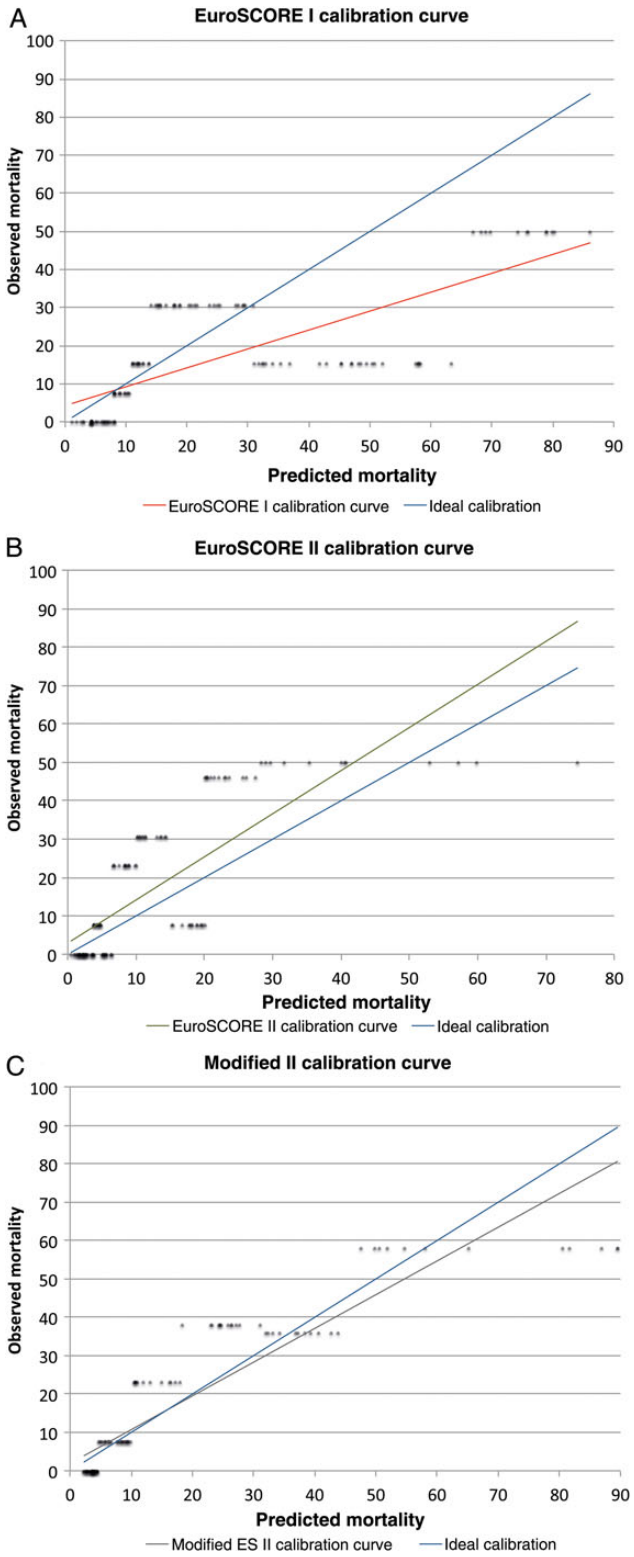

Figure 2: Calibration curves of the EuroSCORE I (**A**), EuroSCORE II (**B**) and modified EuroSCORE II (**C**) predictions. EuroSCORE: European System for Cardiac Operative Risk Evaluation.

DISCUSSION

This single-centre study, based on consecutive patients who underwent cardiac surgery for active IE, demonstrates that both

Table 6: EuroSCORE I versus EuroSCORE II comparison studies

Study	*n*	Population	Discrimination (AUC)		Calibration	
			ES-I	ES-II	ES-I	ES-II
Arnaiz-Garcia *et al.* [20] (Spain)	1200	Mixed	–	–	Over-predicted	Under-predicted
Barili *et al.* [4] (Italy)	12 325	Mixed	0.82	0.82	Over-predicted	Optimal until 30%-predicted mortality than over-predicts
Chalmers *et al.* [15] (UK)	2913	Mixed	0.74	0.82	–	–
Koszta *et al.* [18] (Hungary)	2287	Mixed	0.80	0.82	–	–
Lisboa *et al.* [21] (Brazil)	1000	Mixed	0.81	0.81	HLT, $P = 0.59$	HLT, $P < 0.05$
Nishida *et al.* [23] (Japan)	461	Thoracic aortic surgery	0.72	0.77	Over-predicted	Near optimal
Qadir *et al.* [22] (Pakistan)	2004	Isolated CABG	–	0.84	HLT, $P = 0.23$	HLT, $P < 0.05$
Velicki *et al.* [19] (Serbia)	1247	Mixed	0.76	0.74	Over-predicted	Under-predicted
Wang *et al.* [16] (China)	11 170	Valvular	0.67	0.72	HLT, $P < 0.05$	HLT, $P < 0.05$

AUC: area under the curve; AVR: aortic valve replacement; CABG: coronary artery by-pass grafting; ES: EuroSCORE; HLT: Hosmer–Lemeshow test ($P < 0.05$ indicates poor calibration).

ES-I and ES-II adequately stratify the risk of operative mortality; however, ES-II is better fitted to the current case mix. Also, we found that adding specific IE features to the model would not significantly improve the performance of ES-II, under penalty of increasing model complexity.

It is well known that risk assessment is central in the evaluation of the perioperative risk. The application of risk stratification tools gives an objective appraisal of risk for both physicians and patients. However, some features may not be fully covered by the models, namely (i) centre-to-centre variability in outcomes, (ii) type of surgery required and (iii) inherent complexity of some diseases, in this case IE.

The ES-II and ES-I were conceived from a broad case mix, including few patients with active infective endocarditis and their contribution might have been diluted in the final model. The performance of these scoring systems is unknown in this specific population.

Since its release, the prognostic value of the ES-1 has been extensively proven and its use is recommended as a risk stratification score in both European myocardial revascularization and valvular heart disease guidelines [12, 13]. It was modelled from a surgical population of 1995, wherein most of the procedures were isolated coronary artery bypass grafting (CABG) surgery and only 30% were valvular [5], which is reflected in poorer performance of the model in the later subset of patients [8]. With the improved surgical outcomes and changing demography, the ES-I lost its discrimination and specially calibration power for the nowadays-surgical population [4, 7–9]. The ES-II was conceived to overcome these limitations. It was derived from a population of patients operated in 2010, in which valvular and isolated CABG procedures were well balanced [9]. It showed a very good discriminative ability in both internal and external European validation cohorts (AUC varying between 0.809 and 0.856) [4, 14, 15] and fair discrimination in a Chinese cohort (AUC of 0.72) [16] in overall surgical populations. Several studies demonstrated better discrimination for non-CABG procedures [14–16] though one study revealed an optimal performance in a sub-population of isolated CABG [17]. ES-II performance was lower than expected among patients undergoing emergency/salvage surgery and combined procedures in one study [18]. Despite the improved discrimination, calibration remained unsatisfactory [4, 9, 14–16]. The internal validation study and most of the external validation studies showed a trend toward under-prediction of mortality [9, 14–16, 19], still the ES-II showed optimal calibration until 30% of the predicted risk followed by a progressive over-prediction in high-risk patients in the largest external validation cohort performed by Barili *et al.* [4].

Since the ES-II release in 2011, several comparisons between the updated ES-II and the logarithmic version of the original ES have been made in multiple surgical settings and populations [4, 15, 16, 18–23]. The results are conflicting. The majority of these studies showed an improved performance of the updated version regarding both discrimination and calibration for CABG and non-CABG procedures [15, 16, 18, 23]. However, other reports showed the opposite, being two of these studies from non-European cohorts [19–22]. In the largest comparison ever performed, the ES-II did not seem to significantly improve the performance of the older version in the higher tertiles of risk [4]. Despite these differences, the ES-II performed well (AUC >0.8) in a significant portion of these studies [15, 18, 21, 22].

Table 6 summarizes the main characteristics and results of the aforementioned studies.

In our cohort, the discriminative power of ES-I was less than optimal whereas ES-II showed good discrimination. The AUC of ES-I for perioperative mortality was 0.75, similar to that shown in other nowadays-overall surgical cohorts (AUCs for ES-I between 0.67 and 0.82) [4, 15, 16, 18, 19, 21, 23]. Mestres *et al.* [24] had already validated the ES-I in a registry that included patients with active and past endocarditis; in this study, ES-I showed good calibration and discrimination for the entire cohort and for specific surgical subsets. The ES-II discriminated well with an AUC of 0.83, similar to previous reports from overall surgical cohorts (AUCs for ES-II between 0.72 and 0.84) [4, 14–16, 18, 19, 21–23].

Our study cohort comprises a high-risk subset of patients, which is reflected by the high values of both ES models. In patients within this range of risk, it is expected that the ES-I over-predicts risk [4, 19, 20, 23]; however, such anticipation is not possible for ES-II since its behaviour across the risk spectrum remains controversial [4, 9, 14–16, 19]. Calibration curves were constructed to assess the behaviour of both ES models across the risk spectrum. The ES-I showed poor calibration (Hosmer–Lemeshow test,

$P = 0.04$) and over-predicted mortality. On the other hand, ES-II showed appropriate calibration (Hosmer–Lemeshow test, $P = 0.08$) and a trend to under-prediction. ES-I behaved differently for low-risk and intermediate/high-risk patients. The ES-I showed an initially under-prediction until ~8% of predicted mortality, then intersected the optimal prediction line and exhibited a progressive over-prediction, more pronounced in patients at higher risk. This kind of performance can lead to an unrealistic optimism in patients with less risk and at a preposterous concern in those at highest risk. On the other hand, ES-II showed a progressive under-prediction throughout all risk spectra. Despite these trends for over-prediction in ES-I and under-prediction in ES-II, it appears that both scores are better suited for patients at low risk, which is in consonance with the published literature [4, 8, 9].

For clinical purposes, the ES-II seems to have a more predictable and better behaviour for estimating risk. Overall ES-II performed better than ES-I in predicting perioperative mortality. This was a predictable result, since the ES-II was derived from a population that reflects contemporary surgical practice and demography of most European countries [9], and because the ES-I has worse performance in valvular patients [8].

Several variables not included in the ES models (prosthetic valve IE, perivalvular complications, underlying heart disease and white blood cell count) were associated with perioperative mortality. Only prosthetic valve endocarditis was shown to be a strong independent predictor of mortality (as demonstrated in other series) [1] regardless of ES models. Furthermore, when tested with both ES models separately, ES-I lost its predictive ability. For this reason, we constructed a model including the variables associated with greater IE severity and the ES-II (modified ES-II). The overall performance of the modified ES-II was the best among the three models regarding discrimination and calibration, however did not differ significantly from the ES-II.

In summary, our data suggest that the performance of the ES-I and ES-II in estimating perioperative risk in patients undergoing heart surgery for infective endocarditis is comparable with that of other current overall surgical cohorts, supporting the general trend towards better performance of the ES-II. The greatest advantage of ES-II is its more predictable behaviour across the risk spectrum, allowing more accurate risk estimation. This is of paramount importance for patient and family counselling as well as for surgeon's management of expectations and care. The current model of ES-II is quite comprehensive and additional infective endocarditis features did not increase its performance, reinforcing the evidence that the ES-II is an across-the-board tool for risk estimation in current surgical practice.

Limitations

The inherent limitations of a retrospective analysis are: (i) the subjectivity of those who collect the data and make records (ii) confinement to the data prospectively collected, not allowing other important data to be included. The fact that poor mobility was not discriminated at the time that data were collected implied the assumption that previous serious neurological dysfunction and morbid obesity were surrogates, as well as the description of dependency in the surgeon notes. As all single-centre studies, the external validity is potentially limited. This is particularly important in this setting since microbiological profile, surgical and post-procedural care vary widely between centres. The small sample size (however, it accounts for almost half and one-fifth of the cases included in the ES-I and ES-II development studies, respectively) may have limited the power of the statistical analysis (due to Type II error) and the ability to find statistical significance for many of the comparisons. However, our results are solid and in line with previous evidence on the comparative performance of ES predictive models, and may contribute to further understanding of ES models in this specific surgical subset of patients. Finally, the fact that 55% of the included patients came from secondary centres without cardiac surgery (i) makes it extremely difficult to make assumptions about the time from presentation to diagnosis and from diagnosis to surgery. The later fact may explain why so many patients presented with more than one indication for surgery, suggesting delay in referral (IE is frequently misdiagnosed at presentation and initial management in secondary centres is usually performed by doctors less familiarized with natural history and surgical indications); (ii) may have contributed to a selection bias, with the invoice of patients with more severe and advanced disease but on average younger, with a lower number of comorbidities and at overall lower surgical risk in comparison with the whole population of patients suffering from endocarditis.

CONCLUSIONS

Our analysis of patients with active IE undergoing heart surgery indicates a better overall predictive performance of the EuroSCORE II over the logarithmic EuroSCORE. Specific endocarditis features will increase model complexity without an unequivocal improvement in predictive ability. The present findings indicate that the EuroSCORE II may be a useful and appropriate tool for estimating perioperative risk in a nowadays-active infective endocarditis population. Larger and prospective studies are warranted for further validation.

Conflict of interest: none declared.

REFERENCES

[1] Habib G, Lancellotti P, Antunes MJ, Bongiorni MG, Casalta JP, Del Zotti F *et al.* Guidelines for the management of infective endocarditis: the Task Force for the management of Infective Endocarditis of the European Society of Cardiology (ESC). Endorsed by: European Association for Cardio-Thoracic Surgery (EACTS), the European Association of Nuclear Medicine (EANM). Eur Heart J 2015. DOI: 10.1093/eurheartj/ehv319.

[2] Richardson JV, Karp RB, Kirklin JW, Dismukes WE. Treatment of infective endocarditis: a 10-year comparative analysis. Circulation 1978;58:589–97.

[3] Revilla A, Lopez J, Vilacosta I, Villacorta E, Rollan MJ, Echevarria JR *et al.* Clinical and prognostic profile of patients with infective endocarditis who need urgent surgery. Eur Heart J 2007;28:65–71.

[4] Barili F, Pacini D, Capo A, Rasovic O, Grossi C, Alamanni F *et al.* Does EuroSCORE II perform better than its original versions? A multicentre validation study. Eur Heart J 2013;34:22–9.

[5] 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.

[6] Roques F, Michel P, Goldstone AR, Nashef SA. The logistic EuroSCORE. Eur Heart J 2003;24:881–2.

[7] Lebreton G, Merle S, Inamo J, Hennequin JL, Sanchez B, Rilos Z *et al.* Limitations in the inter-observer reliability of EuroSCORE: what should change in EuroSCORE II? Eur J Cardiothorac Surg 2011;40:1304–8.

[8] Parolari A, Pesce LL, Trezzi M, Cavallotti L, Kassem S, Loardi C *et al.* EuroSCORE performance in valve surgery: a meta-analysis. Ann Thorac Surg 2010;89:787–93, 793.e1–2.

[9] Nashef SA, Roques F, Sharples LD, Nilsson J, Smith C, Goldstone AR *et al.* EuroSCORE II. Eur J Cardiothorac Surg 2012;41:734–44; discussion 744–5.

[10] Li JS, Sexton DJ, Mick N, Nettles R, Fowler VG Jr, Ryan T *et al*. Proposed modifications to the Duke criteria for the diagnosis of infective endocarditis. Clin Infect Dis 2000;30:633–8.
[11] DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 1988;44:837–45.
[12] Kolh P, Wijns W, Danchin N, Di Mario C, Falk V, Folliguet T *et al*. Guidelines on myocardial revascularization. Eur J Cardiothorac Surg 2014; 46:517–92.
[13] Vahanian A, Alfieri O, Andreotti F, Antunes MJ, Baron-Esquivias G, Baumgartner H *et al*. Guidelines on the management of valvular heart disease (version 2012). Eur J Cardiothorac Surg 2012;42:S1–44.
[14] Borracci RA, Rubio M, Celano L, Ingino CA, Allende NG, Ahuad Guerrero RA. Prospective validation of EuroSCORE II in patients undergoing cardiac surgery in Argentinean centres. Interact CardioVasc Thorac Surg 2014;18: 539–43.
[15] Chalmers J, Pullan M, Fabri B, McShane J, Shaw M, Mediratta N *et al*. Validation of EuroSCORE II in a modern cohort of patients undergoing cardiac surgery. Eur J Cardiothorac Surg 2013;43:688–94.
[16] Wang L, Han QQ, Qiao F, Wang C, Zhang XW, Han L *et al*. Performance of EuroSCORE II in patients who have undergone heart valve surgery: a multicentre study in a Chinese population. Eur J Cardiothorac Surg 2014;45: 359–64.
[17] Biancari F, Vasques F, Mikkola R, Martin M, Lahtinen J, Heikkinen J. Validation of EuroSCORE II in patients undergoing coronary artery bypass surgery. Ann Thorac Surg 2012;93:1930–5.
[18] Koszta G, Sira G, Szatmari K, Farkas E, Szerafin T, Fulesdi B. Performance of EuroSCORE II in Hungary: a single-centre validation study. Heart Lung Circ 2014;23:1041–50.
[19] Velicki L, Cemerlic-Adjic N, Pavlovic K, Mihajlovic BB, Bankovic D, Mihajlovic B *et al*. Clinical performance of the EuroSCORE II compared with the previous EuroSCORE iterations. J Thorac Cardiovasc Surg 2014;62: 288–97.
[20] Arnaiz-Garcia ME, Gonzalez-Santos JM, Lopez-Rodriguez J, Dalmau-Sorli MJ, Bueno-Codoner M, Arevalo-Abascal A. Survival after major cardiac surgery: performance and comparison of predictive ability of EuroSCORE II and logistic EuroSCORE in a sample of Mediterranean population. J Thorac Cardiovasc Surg 2014;62:298–307.
[21] Lisboa LA, Mejia OA, Moreira LF, Dallan LA, Pomerantzeff PM, Dallan LR *et al*. EuroSCORE II and the importance of a local model, InsCor and the future SP-SCORE. Rev Bras Cir Cardiovasc 2014;29:1–8.
[22] Qadir I, Alamzaib SM, Ahmad M, Perveen S, Sharif H. EuroSCORE vs. EuroSCORE II vs. Society of Thoracic Surgeons risk algorithm. Asian Cardiovasc Thorac Ann 2014;22:165–71.
[23] Nishida T, Sonoda H, Oishi Y, Tanoue Y, Nakashima A, Shiokawa Y *et al*. The novel EuroSCORE II algorithm predicts the hospital mortality of thoracic aortic surgery in 461 consecutive Japanese patients better than both the original additive and logistic EuroSCORE algorithms. Interact CardioVasc Thorac Surg 2014;18:446–50.
[24] Mestres CA, Castro MA, Bernabeu E, Josa M, Cartana R, Pomar J *et al*. Preoperative risk stratification in infective endocarditis. Does the EuroSCORE model work? Preliminary results. Eur J Cardiothorac Surg 2007;32:281–5.