Heart valve surgery: EuroSCORE vs. EuroSCORE II vs. Society of Thoracic Surgeons score

Muhammad Sharoz Rabbani, Irfan Qadir, Yasir Ahmed, Marrium Gul, Hasanat Sharif

Cardiothoracic Surgery Department, Aga Khan University Hospital, Stadium Road, Karachi - Pakistan

ABSTRACT

Background: This is a validation study comparing the European System for Cardiac Operative Risk Evaluation (EuroSCORE) II with the previous additive (AES) and logistic EuroSCORE (LES) and the Society of Thoracic Surgeons' (STS) risk prediction algorithm, for patients undergoing valve replacement with or without bypass in Pakistan.
Patients and Methods: Clinical data of 576 patients undergoing valve replacement surgery between 2006 and 2013 were retrospectively collected and individual expected risks of death were calculated by all four risk prediction algorithms. Performance of these risk algorithms was evaluated in terms of discrimination and calibration.
Results: There were 28 deaths (4.8%) among 576 patients, which was lower than the predicted mortality of 5.16%, 6.96% and 4.94% by AES, LES and EuroSCORE II but was higher than 2.13% predicted by STS scoring system. For single and double valve replacement procedures, EuroSCORE II was the best predictor of mortality with highest Hosmer and Lemmeshow test (H-L) p value (0.346 to 0.689) and area under the receiver operating characteristic (ROC) curve (0.637 to 0.898). For valve plus concomitant coronary artery bypass grafting (CABG) patients actual mortality was 1.88%. STS calculator came out to be the best predictor of mortality for this subgroup with H-L p value (0.480 to 0.884) and ROC (0.657 to 0.775).
Conclusions: For Pakistani population EuroSCORE II is an accurate predictor for individual operative risk in patients undergoing isolated valve surgery, whereas STS performs better in the valve plus CABG group.

Keywords: Aortic, EuroSCORE, Mitral, Pakistan, Society of Thoracic Surgeons, Valve

Introduction

Risk prediction models play an important role in current cardiac surgical practice. Such models are useful in surgical decision-making, preoperative patient education and consent and quality assurance measures (1). Several tools have been developed in recent years, and among them the Society of Thoracic Surgeons' (STS) mortality risk score and the European System for Cardiac Operative Risk Evaluation (EuroSCORE) scoring system have gained widest popularity (2, 3).

These models have been developed and validated in a surgical population composed mainly of those undergoing coronary artery bypass grafting (CABG). However, valvular heart disease (VHD) comprises a unique subset, with etiologies that are different in developed and developing countries. Most valvular pathology in industrialized countries is now degen-

erative and is a disease of advancing age (4). On the other hand, rheumatic heart disease and infective endocarditis remain endemic in developing countries, inflicting VHD in young patients (5). This difference in clinical profile of patients could affect their postoperative outcomes as well. The objective of this study was to test and compare the predictive performance of additive EuroSCORE (AES), logistic EuroSCORE (LES), EuroSCORE II and STS scoring system in Pakistani patients undergoing valvular surgery.

Patients and Methods

Retrospective data were extracted from the prospectively collected cardiothoracic database of the Aga Khan University Hospital. A total of 576 patients who underwent valvular heart surgery between January 2006 and July 2013 were included. We included mitral valve replacement (MVR), aortic valve replacement (AVR), double valve replacement (DVR), CABG + MVR and CABG + AVR patients. STS calculator does not code for DVR, therefore 490 patients were analyzed for this subgroup. Thirty-day mortality was taken as the primary outcome (hospital mortality or death within 30 days postoperatively). The data set was eventually imported and analyzed using SPSS version 19 software (SPSS, Inc., Chicago, IL, USA).

LES was calculated for each patient based on the original EuroSCORE criteria by running a syntax code incorporating the regression coefficients and intercept developed by EuroSCORE.

Corresponding author:
Hasanat Sharif
Department of Cardiothoracic Surgery
Private wing 2nd floor
Aga Khan University Hospital
Karachi 74800, Pakistan
hasanat.sharif@aku.edu

An individual score for the additive model was also calculated by a simple additive SPSS syntax. The EuroSCORE II and STS risk score were calculated with free online calculators available at http://www.euroscore.org and http://209.220.160.181/STSWebRiskCalc261/de.aspx, respectively. The calculated score corresponds to the individual risk of death in the 30-day post-operative period (predicted mortality).

Patient demographics are presented as percentages for discrete variables and mean (_standard deviation) for continuous variables. Thirty-day mortality was taken as the primary outcome (actual mortality) and determined for the overall patient population, and trends in actual mortality were analyzed across the entire risk spectrum. Performance of the models was also assessed by comparing the observed and expected mortality in quintiles of risk.

The discriminatory power of the models was assessed using the area under the receiver operating characteristic (ROC) curve with 95% confidence intervals; an area of 0.5 indicates no predictive ability, whereas an area of 1.0 represents perfect discrimination (6). Model calibration (the degree to which ob-

served outcomes are similar to the predicted outcomes from the model across patients) of AES, LES, EuroSCORE II and STS 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 test for the lack of "goodness of fit" was applied and graphically represented by a calibration plot. The smooth curve in a calibration plot reflects the nonparametric relationship between observed and predicted risk mortality. The straight dotted line through the origin of a calibration plot represents perfect calibration. Hosmer-Lemeshow p values above 0.05 indicate a well-calibrated model for the study population in question (7).

Results

The study population included 576 patients eligible for calculation of EuroSCORE and 490 patients for STS risk score as it does not stand valid for DVRs. Mean age was 47.36 ± 15.47 years with female population being 46.53%. Table I pres-

TABLE I - Prevalence of risk factors in study population

EuroSCORE II risk factor	AVR	MVR	DVR	CABG + AVR	CABG + MVR
N	137	247	86	49	57
Age					
Mean ± SD	47.98 ± 16.3	43.08 ± 14.17	41.59 ± 13.4	60 ± 11.86	59.8 ± 10.77
Female	29.2	60.3	43	22.4	38.6
Renal					
CC 50-85	43.1	44.9	44.2	45.9	33.3
CC ≤50	10.2	9.7	7	14.3	38.5
Chronic pulmonary disease	2.2	2.8	1.2	2	7
Extra cardiac arteriopathy	1.5	0.4	0	0	0
Insulin-dependent diabetes	3	3.2	1.4	8.2	14
Neurological dysfunction	3	5.3	7	6.2	3.6
Poor mobility	17.5	16.2	10.5	10.2	21.1
Active endocarditis	6.6	4.9	10.5	0	0
Previous cardiac surgery	5.8	13	12.8	12.2	7
Critical preoperative condition	6.6	8.9	4.7	10.2	17.5
Unstable angina	1.5	0	2.3	10.2	8.8
Left ventricular ejection fraction					
Moderate (31-50%)	19.7	12.6	22.1	24.5	38.6
Poor (21-30%)	8	0.4	3.5	6.1	5.3
Very poor (≤20%)	2.9	1.2	0	8.2	14
NYHA (%)					
I	26.3	26.3	26.7	8.2	7
II	35.8	38.1	47.7	42.9	40.4
III	20.4	19.4	15.1	40.8	33.3
IV	17.5	16.2	10.5	8.2	19.3
CCS-4	3.6	4	5.8	16.3	21.1
Recent myocardial infarct	5.1	1.6	2.3	24.5	50.9

AVR = aortic valve replacement; CABG = coronary artery bypass grafting; DVR = double valve replacement; MVR = mitral valve replacement; NYHA = New York Heart Association; CC = creatinine clearance; CCS = Canadian Cardiovascular Society Score.

TABLE II - Performance of risk models in patient subsets

Data set	No. of Patients	Actual	AES			LES			EuroSCORE II			STS		
			Predicted	ROC	H-L p value	Predicted	ROC	H-L p value	Predicted	ROC	H-L p value	Predicted	ROC	H-L p value
All	576	4.86	5.16	0.819	0.184	6.96	0.819	0.233	4.94	0.816	0.552	2.13	0.812	0.290
MVR	247	2.83	4.40	0.881	0.335	5.46	0.877	0.495	3.12	0.898	0.689	1.87	0.882	0.319
AVR	137	2.91	4.44	0.721	0.390	4.93	0.722	0.634	2.49	0.747	0.660	1.29	0.635	0.116
DVR	86	2.32	3.98	0.619	0.431	4.09	0.699	0.371	2.76	0.637	0.346	–	–	–
CABG + MVR	57	1.84	9.30	0.755	0.726	17.30	0.736	0.756	18.22	0.773	0.133	4.84	0.657	0.884
CABG + AVR	49	1.88	8.26	0.587	0.156	13.12	0.581	0.102	9.36	0.521	0.118	2.64	0.775	0.480

AES = additive EuroSCORE; AVR = aortic valve replacement; CABG = coronary artery bypass grafting; DVR = double valve replacement; LES = logistic EuroSCORE; MVR = mitral valve replacement; ROC = receiver operating characteristic; STS = Society of Thoracic Surgeons (STS).
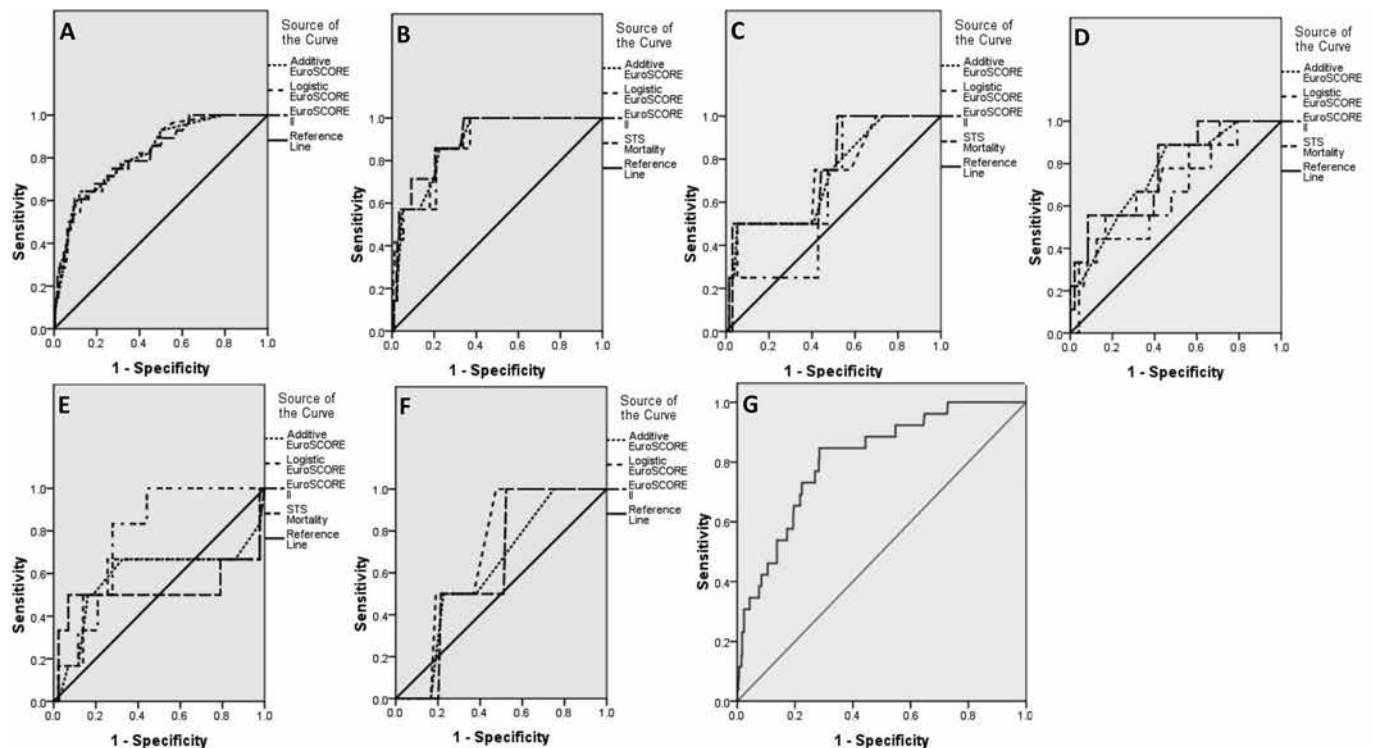

Fig. 1 - Receiver operating characteristic curves. **A)** Entire patient population (576 patients). **B)** MVR patients. **C)** AVR patients. **D)** CABG + MVR patients. **E)** CABG + AVR patients. **F)** DVR patients. **G)** STS ROC for entire patient population (490 patients excluding DVR cases).

ents the distribution of preoperative risk variables identified in our population. Majority of the patients (66.67%) underwent single valve replacement.

There were a total of 28 deaths among 576 operated patients. The actual mortality rate of 4.86% was lower than 5.16%, 6.96% and 4.94% predicted by AES, LES, EuroSCORE II, respectively, but was higher than 2.13% predicted by the STS scoring system. Table II shows the predictive ability, calibration and discriminatory power for all the risk stratifying algo-

rithms. As shown in Figure 1, all the models had good discriminatory power (ROC >0.8) in the global patient sample ranging from 0.812 to 0.819. However, EuroSCORE II was better calibrated with highest Hosmer and Lemmeshow p value (0.552). Table III shows the performance of risk models in quintiles of risk. It shows EuroSCORE II overestimated mortality up to the 5th decile and thereafter under predicted mortality. However, the difference in actual and predicted mortalities was less as compared to differences with AES and LES estimates.

TABLE III - Performance of risk models in quintiles of risk

Risk score	Quintile	All			MVR			AVR			DVR			CABG + MVR			CABG + AVR		
		No. at risk	% observed deaths	% predicted deaths	No. at risk	% observed deaths	% predicted deaths	No. at risk	% observed deaths	% predicted deaths	No. at risk	% observed deaths	% predicted deaths	No. at risk	% observed deaths	% predicted deaths	No. at risk	% observed deaths	% predicted deaths
STS	1st	98	0	0.55	49	0	0.59	27	0	0.44				11	9.09	1.66	10	0	1.27
	2nd	98	2.04	0.84	50	0	0.81	28	0	0.62				12	8.33	2.56	10	0	1.87
	3rd	98	2.04	1.22	49	0	1.02	27	11.1	0.88				11	18.18	3.86	10	10	2.35
	4th	98	7.22	2.07	50	6	1.56	28	0	1.35				12	8.33	5.64	10	30	2.74
	5th	98	15.31	5.98	49	8.16	5.41	27	3.7	3.20				11	36.36	10.64	7	22.2	5.26
AES	1st	115	0	2	56	0	2	36	0	2	21	0	2	18	5.56	6	14	14.29	5.14
	2nd	160	1.25	3	47	0	3	33	3.03	3	16	0	3	9	0	8	8	0	7
	3rd	92	4.35	4.62	47	0	3	34	2.94	4.65	16	6.25	3	9	22.22	9.22	9	0	8
	4th	122	3.28	6.96	44	4.84	4.82	13	0	6	19	5.26	4.95	17	43.53	12.18	10	20	9.50
	5th	87	20.69	11.34	53	7.55	9.06	21	9.52	9.57	14	0	7.86	4	50	15	8	25	13.75
LES	1st	115	0	1.51	55	0	1.51	36	0	1.51	21	0	1.51	11	0	4.65	10	20	3.69
	2nd	128	0.78	2.06	58	0	2.06	21	4.76	2.04	22	0	2.06	12	16.67	8.05	10	0	5.90
	3rd	112	4.46	3.07	10	0	2.25	25	4	2.56	11	9.09	2.27	11	9.09	11.87	11	0	8.74
	4th	106	3.77	6.54	49	4.08	4.10	29	0	4.16	15	0	4.14	12	16.67	23.41	9	11.11	13.54
	5th	115	15.65	22.02	48	10.42	18.09	26	7.69	15.16	17	5.88	11.06	11	36.36	38.81	9	33.33	36.56
EuroSCORE II	1st	115	0	0.61	52	0	0.58	28	0	0.58	18	0	0.97	12	0	3.97	11	18.18	2.76
	2nd	117	1.71	0.88	53	0	0.75	29	0	0.75	16	0	1.23	11	9.09	6.37	9	11.11	4.63
	3rd	115	3.48	1.45	44	0	0.96	52	3.85	1.47	19	5.26	1.59	11	47.27	9.74	10	0	6.65
	4th	115	3.48	3.80	49	4.08	1.73	12	0	3.74	16	6.25	2.57	12	0	22.78	10	0	10.49
	5th	114	15.79	18.16	49	10.20	11.73	16	12.5	11.36	17	0	7.58	11	45.45	49.11	9	33.33	23.94

AES = additive EuroSCORE; AVR = aortic valve replacement; CABG = coronary artery bypass grafting; DVR = double valve replacement; LES = logistic EuroSCORE; MVR = mitral valve replacement; ROC = receiver operating characteristic; STS = Society of Thoracic Surgeons (STS).

For patients undergoing single valve replacement, there was a general trend of overestimation by AES, LES and EuroSCORE II, while STS predicted mortality was lower than the actual value. A total of 14.9% patients had DVR. STS risk score does not code for DVR procedures and was henceforth not used in this comparison. In this subgroup, EuroSCORE II estimates were closest to actual mortality with AES and LES both overestimating mortality. Interestingly the actual mortality paradoxically decreased in patients undergoing DVR as compared to single valve replacement, but the predicted mortality increased as expected. A total of 106 patients underwent mitral valve replacement or AVR along with concomitant CABG surgery, with observed mortality being 1.88% and 1.84%, respectively. Although ROC and H-L p values were lower than those for single valve replacement procedures, when compared with the other three models, STS came out to be the best risk stratifying algorithm for valve replacement surgeries along with bypass.

Discussion

External validation of a risk-predicting model seeks to verify whether it can be generalized outside the boundaries of the population it was built upon, as a preliminary to using it as a reference for quality assessment (8). Differences in the clinical profile of reference population and the testing sample are generally considered as a serious hindrance to such a validation process. In addition, currently used risk score systems have been developed for quite some time and therefore require periodic recalibration and validation to reflect improved surgical techniques and postoperative patient management advances which occurred in recent times (9, 10).

Clinical profile of patients is quite different in Asia as opposed to the Western world. The difference lies in the differences of patient demographics; delayed clinical presentation due to socioeconomic, cultural and geographical reasons; inequitable distribution of medical facilities and different treatment patterns (11). Similar differences combined with altered risk-adjusted mortality were also found in China and India when compared with that of STS and EuroSCORE II sample population (10, 12). The prevalence of rheumatic heart disease (RHD) has fallen drastically in developed countries, with VHD being a disease of the elderly only. On the contrary, prevalence of RHD in Pakistan has remained unchanged from 6 in 1000 in the 1970s to 5.7 in 1000 in 2004 (13). Moreover, Rizvi et al (13) showed that less than 20% of those found to have RHD were aware of their diagnosis before participation in the study and only 8% of affected people (8%) were taking rheumatic prophylaxis. Secondly, we also had a higher incidence of active endocarditis (5.4%) compared to 2.2% in EuroSCORE II population. This may have led to the need for valve replacement at a younger age in our patients. Mean age was 47.4 years as compared to 64.6 years in EuroSCORE II patients (14). About 67.4% of patients in our study were <55 years of age and only 3.8% were above 75 years, compared to 25.6% <55 years and 27.5% above 75 years in the STS database.

Recently the role of EuroSCORE and STS has been questioned. In fact, several papers have suggested that these algorithms may be no more adequate for risk estimation due to a two to threefold overestimation of the adult cardiac surgical risk. The lack of performance has been especially observed in valvular surgical subgroups because these models were developed and validated in a surgical population mainly composed of CABG. EuroSCORE II has been recently published to improve on the older EuroSCORE. In the EuroSCORE II more attention was given on the weight of surgical procedures in order to ameliorate the risk estimation even in non-CABG procedures although the core of risk factors is almost the same (14).

Since publication, the EuroSCORE II has been validated in large, European, multicenter cohorts. Grant et al (15) found that the performance of the tool was "more than satisfactory," and well calibrated for a large proportion of patients. Barili et al (16) demonstrated that the EuroSCORE II is a good predictor of perioperative mortality in isolated aortic valve surgery, with lower discrimination when compared with STS and a better calibration when compared with LES, Age, creatinine and left ventricular ejection fraction model (ACEF) and STS scores. Its performance is optimal in the lowest tertile of patients, whereas it underpredicts mortality afterward. Overall, EuroSCORE II has better performance although it does not seem to significantly improve the performance of older EuroSCORE versions in the higher tertiles of risk. Although high-risk cases constitute only a small proportion of databases from which risk scores are elaborated and validated, they also express a significant proportion of the overall mortality, for which poor approximation in stratification cannot be easily tolerated. In this study, by applying both models on the surgical operation types (Tab. II), no statistically significant differences were observed comparing AUC-AES, AUC-LES with AUC-EuroSCORE II (AUC = Area under the curve). However, EuroSCORE II was better calibrated than the older version. The worst discrimination of all three models was observed in the patients with combined valve and coronary surgery. Similar to our study, Carnero-Alcázar et al also found that EuroSCORE showed excellent discriminative capacity that has been slightly improved with the new EuroSCORE II version (17).

Few studies have drawn head to head comparison of STS and EuroSCORE II (12, 18, 19). Consistent with our findings, Wang et al (19) found that compared with the EuroSCORE II, the STS score gives an accurate prediction for individual operative risk in patients undergoing heart valve surgery with concomitant CABG. Contrary to this, Kirmani et al (20) concluded that EuroSCORE II and STS both provide equivalent discrimination (area under ROC EuroSCORE II vs. STS, 0.816 vs. 0.810, p = 0.714) in predicting mortality in all cardiac surgical procedures. However, STS risk values can be calculated only for selected procedures. In a recent study from India, large proportion of patients undergoing cardiac surgery had to be excluded because STS does not code for DVR, concomitant tricuspid valve surgery or MAZE procedure (12). In our study, we had to exclude 14.9% patients from analysis who underwent surgeries for which STS risk score was not available. Our study showed STS to be well calibrated and showing high discrimination power for CABG and valve replacement patients.

There are several limitations to our study. First, our study is based on a relatively small sample size of patients from a single center, and so the truth of the performance of four risk models for predicting operative mortality in heart valve surgery in Pakistan could be biased, thus the results need to be further

examined using a large number of patients across the multi-center database. Second, this is a retrospective investigation and deviation, such as ignoring a strong predictor (pulmonary hypertension), may weaken the results of the validation procedure. Third, we examined all-cause mortality and were unable to determine the cause of death (cardiac or noncardiac). However, prediction of all-cause mortality is probably more practical than cardiac death alone.

The main results of our study are that (i) EuroSCORE II is the best predictor of mortality in isolated valve surgery patients (including both single valve replacements and DVRs). (ii) STS risk prediction algorithm is better calibrated for CABG + valve replacement. (iii) Clinical profile of Pakistani patients undergoing valvular surgery is uniquely different from European or American population where these risk models were developed and validated.

Conclusion

In conclusion, the EuroSCORE II gives an accurate prediction for individual operative risk in patients undergoing isolated valve surgery, whereas STS performs better in the valve plus CABG group.

Disclosures

Financial support: None of the authors has financial interest related to this study to disclose.
Conflict of interest: The authors declare that they have no conflicting interests.

References

1. Ferraris V, Ferraris S. Risk stratification and comorbidity. Cardiac surgery in the adult, 2nd ed. New York: McGraw-Hill; 2003: 187-224.
2. 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(1):9-13.
3. O'Brien SM, Shahian DM, Filardo G, et al; Society of Thoracic Surgeons Quality Measurement Task Force. The Society of Thoracic Surgeons 2008 cardiac surgery risk models: part 2—isolated valve surgery. Ann Thorac Surg. 2009;88(Suppl 1):S23-S42.
4. Nkomo VT, Gardin JM, Skelton TN, Gottdiener JS, Scott CG, Enriquez-Sarano M. Burden of valvular heart diseases: a population-based study. Lancet. 2006;368(9540):1005-1011.
5. Carapetis JR. Rheumatic heart disease in Asia. Circulation. 2008;118(25):2748-2753.
6. Swets JA. Measuring the accuracy of diagnostic systems. Science. 1988;240(4857):1285-1293.
7. Hosmer DW, Lemeshow S. Applied logistic regression. New York: Wiley; 2004.
8. Zingone B, Pappalardo A, Dreas L. Logistic versus additive EuroSCORE. A comparative assessment of the two models in an independent population sample. Eur J Cardiothorac Surg. 2004; 26(6):1134-1140.
9. Noyez L, Kievit PC, van Swieten HA, de Boer MJ. Cardiac operative risk evaluation: The EuroSCORE II, does it make a real difference? Neth Heart J. 2012;20(12):494-498.
10. Zhang GX, Wang C, Wang L, Lu FL, Li BL, Han L, et al. Validation of EuroSCORE II in Chinese patients undergoing heart valve surgery. Heart Lung Circ. 2013;22(8):606-611.
11. 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. 2011;13(2):137-41.
12. Borde D, Gandhe U, Hargave N, Pandey K, Khullar V. The application of European system for cardiac operative risk evaluation II (EuroSCORE II) and Society of Thoracic Surgeons (STS) risk-score for risk stratification in Indian patients undergoing cardiac surgery. Ann Card Anaesth. 2013;16(3):163-166.
13. Rizvi SF, Khan MA, Kundi A, Marsh DR, Samad A, Pasha O. Status of rheumatic heart disease in rural Pakistan. Heart. 2004;90(4): 394-399.
14. Nashef SA, Roques F, Sharples LD, et al. EuroSCORE II. Eur J Cardiothorac Surg. 2012;41(4):734-744, discussion 744-745.
15. Grant SW, Hickey GL, Dimarakis I, Trivedi U, Bryan A, Treasure T, et al. How does EuroSCORE II perform in UK cardiac surgery; an analysis of 23 740 patients from the Society for Cardiothoracic Surgery in Great Britain and Ireland National Database. Heart. 2012;98(21):1568-1572.
16. Barili F, Pacini D, Rosato F, Parolari A. The role of surgical procedures on discriminative performance of the updated EuroSCORE II. J Thorac Cardiovasc Surg. 2013;146(4):986-987.
17. Carnero-Alcázar M, Silva Guisasola JA, Reguillo Lacruz FJ, et al. Validation of EuroSCORE II on a single-centre 3800 patient cohort. Interact Cardiovasc Thorac Surg. 2013;16(3): 293-300.
18. Barili F, Pacini D, Rosato F, Parolari A. The role of surgical procedures on discriminative performance of the updated EuroSCORE II. J Thorac Cardiovasc Surg. 2013;146(4):986-987.
19. Wang C, Li X, Lu FL, Xu JB, Tang H, Han L, et al. Comparison of six risk scores for in-hospital mortality in Chinese patients undergoing heart valve surgery. Heart Lung Circ. 2013;22(8):612-617.
20. Kirmani BH, Mazhar K, Fabri BM, Pullan DM. Comparison of the EuroSCORE II and Society of Thoracic Surgeons 2008 risk tools. Eur J Cardiothorac Surg. 2013;44(6):999-1005.