ORIGINAL ARTICLE
UDC: 616.12-089-036
https://doi.org/10.2298/VSP170810172N

Clinical performances of EuroSCORE II risk stratification model in the Serbian cardiac surgical population: a single centre validation study including 10,048 patients

Kliničke performanse modela za stratifikaciju operativnog rizika EuroSCORE II kod kardiohirurških bolesnika u Srbiji: studija provere na 10 048 bolesnika operisanih u jednom centru

Duško Nežić*, Tatjana Raguš*, Slobodan Mićović*, Snežana Trajić†, Biljana Spasojević Milin†, Ivana Petrović*, Dragana Košević*, Milorad Borzanović†

Institute for Cardiovascular Diseases "Dedinje", Clinic of Cardiac Surgery, *Department of Cardiac Surgery, †Department of Preoperative Evaluation, Belgrade, Serbia

Abstract

Background/Aim. The EuroSCORE II has recently been developed with an idea to provide better accuracy in prediction of perioperative mortality in the patients who underwent open heart surgery. The aim of this study was to validate clinical performances of the European System for Cardiac Operative Risk Evaluation (EuroSCORE) II risk stratification model in the Serbian adult cardiac surgical population undergoing open heart surgery. **Methods.** The EuroSCORE II values on 10,048 consecutive patients undergoing major adult cardiac surgery from 1st January 2012 to 31st March 2017, were prospectively calculated and entered the institutional database. The discriminative power of the model was tested by calculating the area under the receiver operating characteristic curve (AUC). The calibration of the model was assessed by the Hosmer-Lemeshow (H-L) statistics and the observed to expected (O/E) mortality ratio. The patients with the EuroSCORE II values of 0.5–2.50%, > 2.50–6.50%), and > 6.50% were defined to be at low, moderate, and high perioperative risk, respectively. **Results.** The observed in-hospital mortality was 3.86% (388 of 10,048) and the mean predicted mortality by the EuroSCORE II was 3.61%. The discriminatory power was very good for the entire cohort as well as for all subgroups [coronary, valve(s), combined (coronary plus valve), aortic and other] of performed cardiac procedures (all AUCs > 0.75). The H-L test confirmed good calibration only for category other cardiac procedures. The O/E mortality ratio confirmed good calibration for the whole sample [O/E ratio 1.07, 95% confidence interval (CI) 0.96–1.18] and for all subgroups of performed cardiac procedures, excluding significant underprediction of mortality for aortic surgery (O/E ratio 1.64; 95% CI 1.31–1.97). The EuroSCORE II overestimated perioperative risk in a low and underestimated perioperative risk in a high risk group, with acceptable discrimination (both AUCs = 0.72). On the contrary, the O/E mortality ratio confirmed good calibration for all three subcategories of high risk group. **Conclusion.** The results of our study confirmed acceptable overall performances of the EuroSCORE II risk stratification model in terms of discrimination and the accuracy of model when applied to the contemporary Serbian cardiac surgical cohort undergoing open heart surgery at our Institute.

Key words:
mortality; predictive value of tests; risk assessment; thoracic surgical procedures.

Apstrakt

Uvod/Cilj. EuroSCORE II je razvijen nedavno sa idejom da se obezbedi bolja tačnost u predviđanju perioperativnog mortaliteta bolesnika podvrgnutih operacijama na otvorenom srcu. Cilj rada je bio da se provere kliničke performanse modela za stratifikaciju operativnog rizika u kardiohirurgiji – EuroSCORE II (Evropski sistem za procenu kadiohirurškog operativnog rizika) kod odraslih bolesnika u Srbiji kod kojih se izvode kardiohirurške procedure. **Metode.** Vrednosti EuroSCORE II za 10 048 uzastopno operisanih (od 1. januara 2012. do 31. marta 2017. godine) odraslih kardiohirurških bolesnika prospektivno su izračunate i unete u bazu podataka Instituta za kardiovaskularne bolesti

Correspondence to: Duško Nežić, Institute for Cardiovascular Diseases "Dedinje", Department of Cardiac Surgery, 11 000 Belgrade, Serbia. E-mail: nezic@EUnet.rs

"Dedinje" u Beogradu. Diskriminaciona snaga modela testirana je izračunavanjem površine ispod *receiver operating characteristic* (ROC) krive (AUC). Kalibracija modela je bila procenjena upotrebom Hosmer-Lemeshow (H-L) testa i odnosom između zabeleženog i očekivanog (O/E) mortaliteta. Bolesnici sa vrednostima EuroSCORE II od 0,5% do 2,5% definisani su kao bolesnici sa niskim operativnim rizikom, sa skorom preko 2.5% do 6.5 % sa umerenim, a preko 6,5% sa visokim operativnim rizikom. **Rezultati.** Zabeleženi bolnički mortalitet bio je 3,86% (388 od 10,048), a srednja vrednost mortaliteta predviđenog EuroSCORE-om II iznosila je 3.61%. Diskriminatorna snaga modela je bila vrlo dobra za ceo uzorak, kao i za sve podgrupe [koronarna, valvularna, kombinovana (koronarna plus valvularna) hirurgija, hirurgija grudne aorte i ostalo] izvedenih kardiohirurških procedura (sve AUCs > 0.75). H-L testom potvrđena je dobru kalibracija samo za kategoriju 'druge procedure'. Primenom O/E odnosa potvrđena je dobra kalibracija za ceo uzorak [O/E odnos 1.07, 95% interval pouzdanosti (CI) 0.96–1.18], kao i za sve podgrupe izvedenih kardiohirurških procedura, osim značajnog potcenjivanja mortaliteta u hirurgiji grudne aorte (O/E odnos 1.64; 95% CI 1.31–1.97). EuroSCORE II precenio je operativni rizik u grupi niskog rizika, i potcenio operativni rizik u grupi viskog rizika (O/E odnos mortaliteta), sa prihvatljivom diskriminacijom za obe grupe (AUC = 0.72). Naprotiv, O/E odnos mortaliteta potvrdio je dobru kalibraciju za sve tri potkategorije grupe viskog operativnog rizika. **Zaključak.** Rezultati naše studije potvrđuju prihvatljive opšte performanse (diskriminaciju i kalibraciju) EuroSCORE II modela za stratifikaciju operativnog rizika u kardiohirurgiji, primenjenog na uzorak kardiohirurških bolesnika u Srbiji operisanih u našem Institutu, nakon uvođenja modela u upotrebu.

Ključne reči:
mortalitet; testovi, prognostička vrednost; rizik, procena; hirurgija, torakalna, procedure.

Introduction

Despite the progress in preoperative screening, myocardial protection, surgical techniques, and intensive care unit treatment, the open-heart procedures still carry a risk of mortality and significant morbidity. Risk adjusted perioperative mortality rate following cardiac surgery has been widely used as an indicator of quality of care as well as for comparison of outcomes among institutions and surgeons. In order to assess patients' perioperative risk, several scoring systems were developed over past two decades. Of these risk scores, the Society for Thoracic Surgeons (STS) Predicted Risk of Mortality (PROM) score and the European System for Cardiac Operative Risk Evaluation II (EuroSCORE II) are the most widely used worldwide and they were included in clinical guidelines [1]. The EuroSCORE II [2] has recently been developed as an updated version of the old logistic EuroSCORE with an idea to provide better accuracy (calibration) in prediction of perioperative mortality which aged model overestimated by two- to threefold [2, 3]. Four cardiac centers from Serbia [2] contributed to the EuroSCORE II database (22,381 patients), with more than 1,000 patients [Institute for Cardiovascular Diseases – (ICD) Vojvodina – more than 300 patients [4], ICD Dedinje – almost 500 patients, data for other two centers were approximately calculated]. The initial results of EuroSCORE II validation in the Serbian cardiac cohort were reported by two centers [(ICD Vojvodina [5], and ICD Dedinje [6]), including 1,247 and 1,864 patients (who were operated during 2012), respectively]. The results of coronary artery bypass grafting (CABG), valve(s) and combined [CABG plus valve(s)] surgery were studied in both manuscripts. Although a cohort size of both databases was relatively small, the authors were not able to recruit samples large enough to perform a subanalysis of more specific procedures [i.e. aortic valve replacement (AVR), mitral valve replacement (MVR) or mitral valve repair (MVr) surgery, those procedures combined with CABG, etc]. Therefore, the aim of our study was to validate the EuroSCORE II performances in the contemporary cardiac surgical cohort, large enough to allow more comprehensive analysis of all types of cardiac procedures which were performed during the period of five years.

Methods

The EuroSCORE II values were prospectively calculated using the web-based system (http://www.euroscore.org – this site also include definitions of all EuroSCORE II variables), and .stored in the institutional database for a series of 10,048 consecutive patients who underwent adult (≥ 18 years of age) cardiac surgery at the Institute for Cardiovascular Diseases "Dedinje", Belgrade, Serbia, from 1st January 2012 to 31st March 2017. The patients with a postinfarction ventricular septal rupture were excluded from the study due to a low number of patients with this complication included in the developmental database of EuroSCORE II [2, 7, 8]. Only the first procedure for each patient was entered the registry while any other operation performed during the same in-hospital stay was coded as a complication. The primary end-point for the study was in-hospital mortality (any-cause postoperative death occuring before discharge from the index hospitalization). The patients with the EuroSCORE II values of 0.5%–2.50% > 2.50%–6.50%; and > 6.50% were defined to be at low, moderate, and high perioperative risk, respectively. The high-risk patients were further divided into three categories – higher, very high and extremely high perioperative risk with the EuroSCORE II values of > 6.50%–13.50% > 13.50%–20.00% and > 20.00%, respectively. The Institutional Ethics Committee approved the study and a requirement for informed written consent was waived due to the fact that patients' identities were masked.

The statistical analyses were performed by using the statistical package SPSS version 17.0 (SPSS, Inc., Chicago, IL, USA). The categorical variables were expressed as percentages and continuous variables were expressed as mean ± standard deviation (SD). The statistical analyses were performed by the Fisher's exact test, or the χ^2 test for the cate-

gorical variables and by the *t*-test for the continuous variables. A *p*-value lower than 0.05 was considered significant.

The performance of the EuroSCORE II (for the whole cohort as well as for all subgroups) was analysed focusing on the discrimination power and calibration. Discrimination measures the capacity of the model to distinguish between the patients who will develop an event (in this case perioperative death) and those who will not, namely, to differentiate the low-risk from the high-risk patients. Discrimination can be assessed by the area under the receiver operating characteristic curve (AUC). The AUC is a percentage of randomly drawn pairs (meaning one death and one survivor patient-pairs) for which it is true that a patient who died had a higher risk score than a patient who survived. The discriminative power is thought to be excellent if the AUC is > 0.80, very good if > 0.75 and good (acceptable) if > 0.70 [9].

Calibration refers to the agreement between observed events and predicted probability of occurrence of these events. The Hosmer-Lemeshow (H-L) goodness-of-fit test is the most popular test to validate calibration, measuring the differences between the observed and expected outcomes over deciles of risk. A well-calibrated model gives a corresponding *p*-value > 0.05 [10]. We also evaluated the EuroSCORE II calibration using the observed to expected (O/E) mortality ratio. Ideally, this ratio equals one (the observed mortality equals expected mortality, thus the predictive model is perfectly calibrated). A value above one means that the model underestimates mortality, a value below one means that the model overestimates mortality. If the 95% confidence interval (CI) of the O/E mortality ratio includes the value of 1.0, the model is well calibrated [10].

Results

A total of 10,048 patients were identified to fulfill the study criteria [patients < 18 years of age and patients with postinfarction ventricular septal defect (VSD) were excluded]. The operative details and the patients characteristics (EuroSCORE II variables) of our study population and those of the EuroSCORE II are presented and compared in Table 1. There were no single missing data reffering to the variables necessary for the EuroSCORE models risk calculation. The following subgroups procedures were performed: CABG surgery – 5,228 (52.03%); valve(s) surgery (surgery of one or more valves) – 2,305 (22.94%); combined cases (CABG and valve(s) surgery) – 1,569 (15.62%); aortic (thoracic aorta) surgery – 747 (7.43%) and other major cardiac procedures – 199 (1.98%).

The discriminatory and calibration abilities of EuroSCORE II pulled out from the initial studies of the Serbian cardiovascular centers [5, 6] are summarised in Table 2, as well as the ICD "Dedinje" data including over 10,000 patients. The in-hospital mortality observed in our 5-year sample was 3.86% (388 out of 10,048), while the EuroSCORE II predicted mortality of 3.61%. For that period, the EuroSCORE II showed very good discriminative power in all categories (all AUCs > 0.75); for the whole cohort and for all subgroups procedures which were performed – CABG, valve(s), combined, aortic, others) (Table 2). The EuroSCORE II discriminative power was also good for almost all (12 out of 14) more specific procedures [aortic valve surgery, mitral valve surgery, multiple valve surgery as well as for those procedures combined with CABG surgery, excluding MVR plus tricuspid valve surgery (TVs) and MVr plus TVs plus CABG) (Table 3). Although the H-L statistics failed to confirm the overall and subgroups good calibration (except in the category 'others': the H-L *p* value of 0.61), it confirmed a good calibration of EuroSCORE II model for all more specific procedures (Table 3). On the contrary, the O/E mortality ratio confirmed good calibration for the whole sample and for all subgroups of performed cardiac procedures, excluding aortic surgery (a significant underestimation of mortality; O/E mortality ratio = 1.64; 95% CI: 1.31–1.97) (Table 2). The O/E mortality ratio confirmed a good calibration for all more specific procedures (aortic valve surgery, mitral valve surgery, multiple valve surgery, and those procedures combined with CABG surgery), except for MVR plus TVs plus CABG, showing a significant overprediction of mortality by the EuroSCORE II (O/E ratio 0.40; 95% CI: -0.15–0.95) (Table 3). The EuroSCORE II discriminative power was acceptable (AUCs = 0.72) for the low-risk and high-risk groups while it failed to confirm a good discrimination in the moderate-risk group as well as in all subcategories of high-risk group (Table 4). The H-L statistics failed to confirm a good calibration only for the high-risk group. The O/E mortality ratio confirmed a good calibration for the moderate-risk group and for all subcategories of high-risk group. On the contrary, for the high-risk category, the O/E mortality ratio showed a significant overprediction of mortality (O/E mortality ratio = 1.24; 95% CI: 1.08–1.40). However, further analysis of high risk group subcategories confirmed good calibration for all subcategories (O/E mortality ratio and H-L statistics) (Table 4). For the low-risk group, the O/E mortality ratio showed a significant underestimation of mortality (O/E mortality ratio = 0.66; 95% CI: 0.48–0.84).

Discussion

The significant progress in the development of risk prediction models in cardiac surgery was made in the last two decades. Therefore, a risk- adjusted perioperative mortality rate following cardiac surgery was widely used as an indicator for a quality of care as well as for comparison of outcomes among institutions and surgeons. Predicted probability of occurrence of postoperative death also enabled stratification of patients into the different clinical risk groups (low, moderate, high), and, subsequently, made it possible to target the high-risk surgical patients in need of new therapeutic interventions [11]. We have to point out that there is no ideal cardiac surgical risk prediction score model available. Although the Society of Thoracic Surgeons Predicted the Risk of Mortality (STS PROM) score (including 40 variables) and the EuroSCORE II (including 17 variables) were integrated into clinical guidelines [1], they are still missing some risk factors claimed to significantly contribute to perioperative mortality in cardiac surgery (preoperative anemia [12, 13], liver dysfunction [14, 15], and frailty [16, 17]).

Statistical explanation for omission of these risk factors might be that factors with a low prevalence, even if associated with a high odds ratio for the outcome (in this case – perioperative mortality) at the univariate analysis, are generally excluded by the multivariable logistic regression models [12]. Although aged, the EuroSCORE models (additive and logistic) retained a good discriminative power over the time; they failed to maintain good calibration due to an overestimation of the adult cardiac surgical risk by two-to threefold [2, 3, 18]. Therefore, they were updated and renewed into the EuroSCORE II (EuroSCORE Pilot Study, 2010). The internal testing of EuroSCORE II performances on the validation dataset (5,553 patients) confirmed a good discrimination [2] and a good calibration, too (H-L test, $p = 0.09$) [19]. The EuroSCORE II performances underwent an external validation in numerous studies, later on. Grant et al. [20] presented a validation of EuroSCORE II in a sample of 23,740 patients and supported the use of EuroSCORE II as a generic risk model for the United Kingdom contemporary cardiac surgery. Garcia-Valentin et al. [21] in their prospective, multicentre study (20 centers, 4,034 patients) concluded that the EuroSCORE II was the best model in Spain at that moment.

Table 1

Patients characteristics and operative details for the study population compared with the original EuroSCORE II dataset

Variable	EuroSCORE II		*p*-value
	(our database, n = 10,048)	(original database n = 22,381)*	
Age (years), mean ± SD	63.2 ± 10.5	64.6 ± 12.5	0.0001
Gender (female), n (%)	2,963 (29.5)	6,919 (30.9)	0.01
Renal impairment, n (%)			
normal	5,218 (51.9)		
moderate	3,826 (38.1)		
severe	951 (9.5)		
dialysis	53 (0.5)	244 (1.1)	0.0001
Extracardiac arteriopaty, n (%)	1,769 (17.6)		
Poor mobility, n (%)	71 (0.7)		
Previous cardiac surgery, n (%)	359 (3.6)		
Chronic lung disease, n (%)	529 (5.3)	2,384 (10.7)	0.0001
Active endocarditis, n (%)	122 (1.2)	497 (2.2)	0.0001
Critical preoperative care, n (%)	91 (0.9)	924 (4.1)	0.0001
Diabetic on insulin, n (%)	1,028 (10.2)	1,705 (7.6)	0.0001
NYHA Class, n (%)			
I	1,331 (13.2)		
II	6,141 (61.1)		
III	2,440 (24.3)		
IV	136 (1.4)		
CCS Class IV, n (%)	797 (7.9)		
Left ventricle function, n (%)			
good	3,900 (38.8)		
moderate	4,632 (46.1)		
poor	965 (9.6)		
very poor	551 (5.5)		
Recent myocardial infarction, n (%)	1,209 (12.0)		
Pulmonary hypertension, n (%)			
moderate	2,932 (29.2)		
severe	881 (8.8)		
Urgency, n (%)			
elective	7,590 (75.5)	17,165 (76.7)	0.02
urgent	1,763 (17.5)	4,135 (18.5)	0.04
emergency	684 (6.1)	972 (4.3)	0.0001
salvage	11 (0.1)	109 (0.5)	0.0001
Weight of the intervention, n (%)			
isolated CABG	5,228 (52.0)	10,448 (46.7)	0.0001
single non-CABG	2,002 (19.9)		
two procedures	2,007 (20.0)		
three procedures	811 (8.1)		
Surgery on thoracic aorta, n (%)	747 (7.4)	1,636 (7.3)	0.70
EuroSCORE II (%)	3. 61	3. 90	

***Data available from the original manuscript by Nashef et al. [2]; EuroSCORE – European System for Cardiac Operative Risk Evaluation; NYHA – New York Heart Association; CCS – Canadian Cardiovascular Society; CABG – coronary artery bypass grafting.**

n (%) – number (percentage) of patients.

Table 2

Validation of EuroSCORE II performances in Serbian cardiovascular centers

Center Cardiac procedures (number of patients)	Mortality, n (%)		O/E ratio	(95% CI)	H-L *p*-value	AUC (95% CI)
	observed	expected				
ICD Vojvodina [5]						
All patients (n = 1,247)	43 (3.45)	27 (2.13)	1.59	(1.12–.06)	0.14	0.74 (0.67–0.82)
CABG (n = 718)	16 (2.23)	12 (1.67)	1.33	(0.68–1.98)	0.035	0.72 (0.58–0.86)
Valve(s) (n = 294)	11 (3.74)	6 (2.00)	1.83	(0.7–2.91)	0.49	0.73 (0.57–0.89)
Combined (n = 233)	16 (6.87)	9 (3.65)	1.78	(0.91 –2.65)	0.64	0.68 (0.53–0.82)
Aortic /						
Other CP /						
ICD Dedinje [6], (one-year data)						
All patients (n = 1,864)	68 (3.65)	65 (3.51)	1.05	(0.81–1.29)	0.003	0.85 (0.81–0.89)
CABG (n = 1,039)	24 (2.31)	25 (2.39)	0.96	(0.58–1.34)	0.038	0.81 (0.72–0.91)
Valve(s) (n = 410)	15 (3.66)	14 (3.42)	1.07	(0.5–1.61)	0.26	0.91 (0.86–0.96)
Combined (n = 260)	13 (5.00)	16 (6.16)	0.81	(0.37–1.25)	0.52	0.72 (0.58–0.86)
Aortic (n = 122)	16 (13.1)	10 (8.00)	1.60	(0.82–2.38)	0.28	0.82 (0.74–0.91)
Other CP (n = 33)	0 (0.00)	1 (2.47)	N/A	N/A	N/A	N/A
ICD Dedinje > 5 year data,						
(> 10,000 patients)						
All patients (n = 10,048)	388 (3.86)	363 (3.61)	1.07	(0.96 – 1.18)	0.0001	0.84 (0.82–0.86)
CABG (n = 5,228)	117 (2.24)	126 (2.41)	0.93	(0.76 – 1.10)	0.0001	0.84 (0.80–0.87)
Valve(s) (n = 2,305)	69 (2.99)	71 (3.10)	0.97	(0.74 – 1.20)	0.006	0.86 (0.81–0.90)
Combined (n = 1,569)	99 (6.31)	102 (6.51)	0.97	(0.78 – 1.26)	0.001	0.78 (0.73–0.83)
Aortic (n = 747)	95 (12.7)	58 (7.82)	1.64	(1.31 – 1.97)	0.005	0.76 (0.70–0.81)
Other CP (n = 199)	8 (4.02)	5 (2.60)	1.60	(0.49 – 2.71)	0.61	0.79 (0.61–0.98)

EuroSCORE – European System for Cardiac Operative Risk Evaluation; ICD – Institute for Cardiovascular Diseases; N/A – not applicable; CP – cardiac procedures; CABG – coronary artery bypass grafting; O/E – observed to expected; H-L – Hosmer-Lemeshow; AUC – area under the receiver operating characteristic curve.

Table 3

Calibration and discrimination of EuroSCORE II across more specific procedures [valve(s) and combined surgery]

Type of surgery	Patients (n)	Observed mortality n (%)	Expected mortality n (%)	O / E ratio	(95% CI)	H-L *p*-value	AUC (95% CI)
AVR	(979)	21 (2.15)	24 (2.44)	0.88	(0.51–1.25)	0.18	0.86 (0.78–0.93)
AVR + CABG	(835)	44 (5.27)	46 (5.51)	0.96	(0.68–1.24)	0.55	0.74 (0.66–0.82)
MVR	(399)	11 (2.76)	12 (2.96)	0.92	(0.38–1.46)	0.44	0.81 (0.70–0.92)
MVR + CABG	(217)	20 (9.22)	16 (7.18)	1.25	(0.70–1.80)	0.14	0.82 (0.72–0.92)
MVR + TVs	(157)	7 (4.46)	8 (5.11)	0.88	(0.23–1.53)	0.37	0.64 (0.40–0.88)
MVR + TVs + CABG	(45)	2 (4.44)	5 (11.05)	0.40	(-0.15–0.95)	0.99	0.95 (0.00–1.00)
MVr	(437)	3 (0.69)	6 (1.46)	0.50	(-0.07–1.07)	0.51	0.70 (0.00–1.00)
MVr + CABG	(286)	13 (4.55)	16 (5.51)	0.81	(0.37–1.25)	0.23	0.85 (0.37–1.25)
MVr + TVs	(85)	5 (5.88)	3 (3.53)	1.67	(0.21–3.13)	0.80	0.83 (0.68–0.98)
MVr + TVs + CABG	(45)	3 (6.67)	5 (10.81)	0.60	(-0.08–1.28)	0.36	0.45 (0.00–0.92)
AVR + MVR	(157)	14 (8.92)	11 (7.10)	1.27	(0.60–1.94)	0.49	0.84 (0.73–0.96)
AVR + MVR + CABG	(64)	8 (12.5)	7 (10.7)	1.14	(0.35–1.93)	0.38	0.88 (0.00–1.00)
AVR + MVr	(91)	8 (8.79)	7 (8.02)	1.14	(0.35–1.93)	0.34	0.89 (0.78–1.00)
AVR + MVr + CABG	(77)	9 (11.7)	8 (10.44)	1.13	(0.40–1.86)	0.66	0.78 (0.59–0.96)

EuroSCORE – European System for Cardiac Operative Risk Evaluation; AVR – aortic valve replacement; MVR – mitral valve replacement; MVr – mitral valve reconstruction; TVs – tricuspid valve surgery; CABG – coronary artery bypass grafting; O/E – observed to expected; CI – confidence interval; H-L – Hosmer-Lemeshow; AUC – area under the receiver operating characteristic curve; n – number of patients.

Table 4

Calibration and discrimination of EuroSCORE II across arbitrary determined risk group categories

EuroSCORE II risk group (predicted risk %)	Mortality, n (%)		O/E ratio	(95% CI)	H-L *p*-value	AUC (95% CI)
	observed	expected				
Low (0.5–2.5) [6,000 (59.7 %)]	52 (0.87)	79 (1.32)	0.66	(0.48–0.84)	0.81	0.72 (0.65–0.79)
Moderate (> 2.5–6.5) [2,730 (27.2%)]	118 (4.32)	108 (3.96)	1.09	(0.89–1.29)	0.18	0.64 (0.58–0.69)
High (> 6.5) [1,318 (13.1%)]	218 (16.5)	176 (13.4)	1.24	(1.08–1.40)	0.007	0.72 (0.68–0.75)
Higher (> 6.5–13.5) [923 (9.18%)]	103 (11.16)	84 (9.08)	1.23	(0.99–1.47)	0.78	0.67 (0.61–0.759
Very high (> 13.5–20.0) [216 (2.14%)]	46 (21.3)	35 (16.3)	1.31	(0.93–1.69)	0.22	0.51 (041–0.60)
Extremely high (> 20.0) [179 (1.78%)]	69 (38.55)	57 (31.8)	1.21	(0.92–1.50)	0.38	0.68 (0.60–0.76)

EuroSCORE – European System for Cardiac Operative Risk Evaluation; O/E – observed to expected; CI – confidence interval; H-L – Hosmer-Lemeshow; AUC – area under the receiver operating characteristic curve.

In a meta-analysis [22] of 22 studies involving 145,592 cardiac surgery procedures, the authors concluded that the EuroSCORE II showed good overall performances in terms of discrimination and accuracy of model prediction for operative mortality in cardiac surgery. Although four cardiac centers from Serbia [2] contributed to the EuroSCORE II development dataset with more than 1,000 patients (≈ 5% of database), it would be reasonable to expect that the EuroSCORE II would also be an appropriate model for prediction of operative mortality in Serbian patients undergoing open heart surgery. Indeed, the initial results of EuroSCORE II validation in the Serbian cardiac surgical cohort [5, 6], confirmed an overall good discriminative power. Calibration using the O/E mortality ratio was good in all categories, excluding a significant underprediction of mortality (O/E ratio 1.59; 95% CI 1.12–2.06) for the category 'all patients' of ICD Vojvodina sample [5] (Table 2).

The analysis of our 5-year results confirmed a very good discriminative power of the EuroSCORE II for the whole cohort (AUC = 0.84) as well as for all subgroups of performed cardiac procedures (AUCs from 0.76 to 0.86) (Table 2). The H-L statistics confirmed a good calibration only for the subgroup 'other cardiac procedures'. Although we tested a huge sample, it was not a big surprise that the H-L goodness of fit test did not perform well. Namely, in order to achieve proper conditions to obtain more precise calibration, Paul et al. [23] requested that the huge samples should be divided into more groups (> 2,000 patients in 34, and > 4,000 patients in 130 groups), which was impossible to perform using the statistical package SPSS version 17.0. On the contrary, the O/E mortality ratio (including 95% CI values) confirmed a good calibration for all categories, except for the 'aortic surgery' (significant underestimation of mortality, O/E ratio 1.64; 95% CI: 1.31–1.97). Although O/E mortality ratio for aortic surgery was very close to our previously published [6] result (1.64 vs. 1.60) with a larger sample, the difference (showing underestimation of mortality) became statistically significant. So far, a very few authors have reported results on the thoracic aorta surgery using the EuroSCORE II prediction of mortality. Chalmers et al. [24] reported in-hospital mortality of 6.8% with the median EuroSCORE II value of 5.6% (interquartile range 3.1% to 11.1%). Nishida et al. [25] presented 461 consecutive patients undergoing thoracic aorta surgery with the observed mortality of 7.2%, with the average EuroSCORE II value of 7.4%. We have to point out the possibility that, generally speaking, the overestimation of mortality risk by the risk stratification model may result from publication bias, namely, studies which obtained favorable results could be reported more easily, while authors with unfavorable results (significantly worse outcome compared with predicted mortality) are not so willing to publish their results [22, 26, 27]. Therefore, we have checked our results in the elective and urgent/emergent aortic surgery. In the elective aortic surgery [mortality 5.07% (22 out of 434)], a discriminative power of EuroSCORE II was acceptable (AUC = 0.702), while calibration showed a nonsignificant overprediction of mortality (expected mortality of 6.49%; O/E ratio 0.79; 95% CI 0.46–1.12) by the EuroSCORE II. Thus, for the elective aortic surgery, the EuroSCORE II confirmed a good discrimination and calibration. In the urgent/emergent aortic surgery [mortality 23.3% (73 out of 313)] a discriminative power of the EuroSCORE II was good (AUC = 0.74), while calibration showed a significant underprediction of mortality (expected mortality of 9.7%; O/E ratio 2.43; 95% CI 1.87–2.99) by the EuroSCORE II. However, in a real world scenario, in the patients undergoing aortic surgery for the acute aortic dissection, early mortality still remains high, ranging from 17% to 26% [28–32]. Underprediction of operative mortality by the EuroSCORE II in this category might be attributed to the fact that some very important risk factors are not included in the EuroSCORE II variables [neurological dysfunction/comma, organ system malperfusion (especially visceral/mesenteric ischemia/infarction), hypotension, possible cardiac tamponade, ongoing cardiac ischemia, etc] [29, 30, 32]. Currently, the observations from the German Registry for the Acute Aortic Dissection Type A (GERAADA) (50 cardiac surgery centers in Austria, Switzerland and Germany, including 2,137 patients), confirm a progressively escalating

mortality with each additional malperfused organ system (adjusted odds ratio for one organ = 1.65, two organs = 2.44, three, or more = 3.39; $p < 0.0001$) [32].

The overall observed mortality for our whole cohort showed a slight, nonsignificant underprediction of mortality (O/E ratio 1.07; 95% CI 0.96–1.18). Several studies, including thousands of patients, confirmed that the EuroSCORE II significantly overpredict mortality [Guida et al. [22] (145,592 patients), O/E = 0.89, 95% CI 0.86–0.92; Grant et al. [20] (23,740 patients), O/E = 0.91, 95% CI 0.84–0.98; Osnabrugge et al. [33] [21,016 patients, operated after 1st January 2008; it is a part of a whole cohort (50,558 patients) which represent more contemporary sample, and it would be used in further comparisons, O/E = 0.80, 95% CI 0.73–0.87]. On the other hand, several studies confirmed a significant underprediction of mortality by the EuroSCORE II [Velicki et al. [5] (1,247 patients), O/E = 1.59, 95% CI 1.12–2.06; Arnaiz-Garcia et al. [34] (1,200 patients), O/E = 1.86, 95% CI 1.46–2.26; Chalmers et al. [24] (5,576 patients), O/E = 1.71, 95% CI 1.47–1.95 [35]]. In the CABG, valve(s) and combined surgery in our cohort, the observed mortality was slightly (but not statistically significantly) better than the predicted by the EuroSCORE II (Table 2). The significantly better results (compared with predicted mortality by EuroSCORE II) in CABG surgery were reported by Grant et al. [20] (12,470 patients), O/E = 0.71, 95% CI 0.61–0.81; as well as by Osnabrugge et al. [33] (16,096 patients), O/E = 0.77, 95% CI 0.74–0.80]. On the contrary, Kunt et al. [18] (428 CABG patients over 70 years of age) presented extremely poor prediction, with an O/E ratio of 4.86, 95% CI 3.03–6.43. To the best of our knowledge, the EuroSCORE II validation of more specific procedures (AVR, MVR, MVr, multiple valve surgery, and combined procedures – valve(s) surgery with CABG surgery), have not yet been presented for the Serbian cardiac surgical population. For AVR and AVR plus CABG surgery, we performed slightly (not significantly) better than predicted by the EuroSCORE II (O/E mortality ratio of 0.88 and 0.96, respectively). Those results coincides with large series by Osnabrugge et al. [33] (2,170 AVR patients, and 1,627 AVR plus CABG cases, O/E ratio of 1.14, 95% CI 0.87–1.40 and 0.76, 95% CI 0.58–0.95; respectively), and by Biancari et al. [36] including 11,791 AVR patients, with the O/E ratio of 0.94. In mitral valve reconstructive surgery, we achieved excellent result (mortality 0.69%, 3 of 437 patients) compared with the predicted mortality (1.46%) by the EuroSCORE II (O/E ratio of 0.5, 95% CI -0.07–1.07). This result is almost comparable with reference mitral valve center (Ottawa) [37] result – mortality of 0.60% (5 of 851, O/E ratio of 0.24, 95% CI 0.03–0.51). For MVr surgery, Osnabrugge et al. [33] reported 624 patients with the O/E ratio of 0.64, 95% CI 0.17–1.11. In MVR surgery, Chan et al. [37] again presented the significantly better result than predicted by the EuroSCORE II (6 of 303, O/E ratio of 0.44, 95% CI 0.07–0.81). On the contrary, the Osnabrugge's group [33] observed nonsignificant underestimation of mortality by the EuroSCORE II for MVR surgery (O/E ratio 1.34, 95% CI 0.87–1.81). We (in 399 patients) observed slightly better results than predicted (O/E ratio 0.92). For all other more specific procedures, usable data are not available in relevant literature (small samples, incomplete data, non-contemporary cohorts, etc).

The acceptable discriminative power of EuroSCORE II was detected for low-risk (AUC – 0.72) and high-risk group (AUC – 0.72) category. The EuroSCORE II failed to confirm a good discriminative power for moderate-risk category and in all high-risk group subcategories. The explanation for reduced discriminative power is statistically simple. When the patients are stratified according to the risk score, and then only one strata is analysed, the regressors and their coefficients within the stratum are different from those which allocated them to that risk group in the first place [38]. Furthermore, a minimum of 100 (and preferably 200) events (perioperative deaths) should be included in the sample size so that the model performance can be adequately assessed [39]. The HL statistics failed to confirm a good calibration only for the high-risk group category. According to the O/E mortality ratio, for the low-risk group, the model significantly overestimate mortality (O/E ratio 0.66, 95% CI 0.48–0.84), while it slightly, but significantly underestimate mortality in the high-risk group (O/E 1.24, 95% CI 1.08–1.40). On the contrary, further analysis of high-risk group subcategories confirmed a good calibration for all subcategories. Although we confirmed a good calibration in all subcategories of high-risk category, our results are not in accordance with the previous statements that the EuroSCORE II significantly underestimate mortality in the high-risk group category [5, 8, 40, 41]. Regarding the results in all high-risk group subcategories, our study is in keeping with the results of Barili et al. [3] who showed an optimal EuroSCORE II calibration until 30%-predicted mortality. The results of our study show an acceptable overall performances of the EuroSCORE II risk stratification model in terms of discrimination and accuracy of model, when applied to the Serbian contemporary cardiac surgical cohort undergoing open heart surgery at our Institute.

Limitations of the study

The limitation of our study is its single-center design, and therefore results may not represent national and international practice and outcomes. Although our cohort recruited more than 10,000 patients, another limitation was a sample size, which generated relatively small specimens, including the limited number of tested events (in this case perioperative deaths) for more precise analysis of some subgroups.

Conclusion

The results of our study confirmed acceptable overall performances of the EuroSCORE II risk stratification model in terms of discrimination and accuracy of model, when applied to the Serbian contemporary cardiac surgical cohort undergoing open heart surgery at our Institute.

Conflict of interest

None declared.

REFERENCES

1. *Sullivan PG, Wallach JD, Ioannidis JP.* Meta-Analysis Comparing Established Risk Prediction Models (EuroSCORE II, STS Score, and ACEF Score) for Perioperative Mortality During Cardiac Surgery. Am J Cardiol 2016; 118(10): 1574–82.
2. *Nashef SA, Roques F, Sharples LD, Nilsson J, Smith C, Goldstone AR,* et al. EuroSCORE II. Eur J Cardiothorac Surg 2012; 41(4): 734–44; discussion 744–5.
3. *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(1): 22–9.
4. *Mihajlovic B.* Risk stratification and prediction of operative treatment outcome in cardiosurgery. Med Pregl 2011; 64(3–4): 133–6.
5. *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. Thorac Cardiovasc Surg 2014; 62(4): 288–97.
6. *Nezic D, Spasic T, Micovic S, Kosevic D, Petrovic I, Lausevic-Vuk L,* et al. Consecutive Observational Study to Validate EuroSCORE II Performances on a Single-Center, Contemporary Cardiac Surgical Cohort. J Cardiothorac Vasc Anesth 2016; 30(2): 345–51.
7. *Noyez L, Kievit PC, Swieten HA, Boer MJ.* Cardiac operative risk evaluation: The EuroSCORE II, does it make a real difference. Neth Heart J 2012; 20(12): 494–8.
8. *Di Dedda U, Pelissero G, Agnelli B, De Vincentiis C, Castelvecchio S, Ranucci M.* Accuracy, calibration and clinical performance of the new EuroSCORE II risk stratification system. Eur J Cardiothorac Surg 2013; 43(1): 27–32.
9. *Nashef SA.* Death and quality in cardiac surgery. J Patient Safety Risk Manag 2010; 16(4): 130–4.
10. *Nezic D, Borzanovic M, Spasic T, Vukovic P.* Calibration of the EuroSCORE II risk stratification model: Is the Hosmer-Lemeshow test acceptable any more. Eur J Cardiothoracic Surg 2013; 43(1): 206.
11. *Papadopoulos N, Wenzel R, Thudt M, Doss M, Wimmer-Greinecker G, Seeger F,* et al. A Decade of Transapical Aortic Valve Implantation. Ann Thorac Surg 2016; 102(3): 759–65.
12. *Ranucci M, Di Dedda U, Castelvecchio S, Menicanti L, Frigiola A, Pelissero G.* Surgical and Clinical Outcome Research (SCORE) group. Impact of preoperative anemia in outcome in adult cardiac surgery: A propensity-matched analysis. Ann Thorac Surg 2012; 94(4): 1134–42.
13. *Scrascia G, Guida P, Caparrotti SM, Capone G, Contini M, Cassese M,* et al. Incremental value of anemia in cardiac surgical risk prediction with the European System for Cardiac Operative Risk Evaluation (EuroSCORE) II model. Ann Thorac Surg 2014; 98(3): 869–75.
14. *Hsieh WC, Chen PC, Corciova F, Tinica G.* Liver dysfunction as an important predicting risk factor in patients undergoing cardiac surgery: A systematic review and meta-analysis. Int J Clin Exp Med 2015; 8(11): 20712–21.
15. *Diaz GC, Renz JF.* Cardiac surgery in patients with end-stage liver disease. J Cardiothorac Vasc Anesth 2014; 28(1): 155–62.
16. *Abdullahi YS, Athanasopoulos LV, Casula RP, Moscarelli M, Bagnall M, Ashrafian H,* et al. Systematic review on the predictive ability of frailty assessment measures in cardiac surgery. Interact Cardiovasc Thorac Surg 2017; 24(4): 619–24.
17. *Sündermann SH, Dademasch A, Seifert B, Rodriguez CB, Emmert MY, Walther T,* et al. Frailty is a predictor of short- and mid-term mortality after elective cardiac surgery independently of age. Interact Cardiovasc Thorac Surg 2014; 18(5): 580–5.
18. *Kunt AG, Kurtcephe M, Hidiroglu M, Cetin L, Kucuker A, Bakuy V,* et al. Comparison of original EuroSCORE, EuroSCORE II and STS risk models in a Turkish cardiac surgical cohort. Interact Cardiovasc Thorac Surg 2013; 16(5): 625–9.
19. *Sharples LD, Nashef SA.* EuroSCORE Project Group. Reply to Nezic et al. Eur J Cardiothorac Surg 2013; 43: 207.
20. *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): 1568–72.
21. *Garcia-Valentin A, Mestres C, Bernabeu E, Bahamonde JA, Martín I, Rueda C,* et al. Validation and quality measurements for EuroSCORE and EuroSCORE II in the Spanish cardiac surgical population: A prospective, multicentre study. Eur J Cardiothoracic Surg 2016; 49: 399–405.
22. *Guida P, Mastro F, Scrascia G, Whitlock R, Paparella D.* Performance of the European System for Cardiac Operative Risk Evaluation II: A meta-analysis of 22 studies involving 145,592 cardiac surgery procedures. J Thorac Cardiovasc Surg 2014; 148(6): 3049–57.e1.
23. *Paul P, Pennell ML, Lemeshow S.* Standardizing the power of the Hosmer: Lemeshow goodness of fit test in large data sets. Stat Med 2013; 32(1): 67–80.
24. *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 Cardithorac Surg 2013; 43: 688–94.
25. *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 aorta surgery in 461 consecutive Japanese patients better than both the original additive and logistic EuroSCORE algorithms. Interact Cardiovasc Thorac Surg 2014; 14: 446–50.
26. *Siregar S, Groenwold RH, de Heer F, Bots ML, van der Graaf Y, van Herwerden LA.* Performance of the original EuroSCORE. Eur J Cardiothorac Surg 2012; 41(4): 746–54.
27. *Ranucci M, Castelvecchio S, Menicanti LA, Scolletta S, Biagioli B, Giomarelli P.* An adjusted EuroSCORE model for high-risk cardiac patients. Eur J Cardiothorac Surg 2009; 36(5): 791–7.
28. *Erbel R, Aboyans V, Boileau C, Bossone E, Bartolomeo RD, Eggebrecht H,* et al. ESC Committee for Practice Guidelines. 2014 ESC Guidelines on the diagnosis and treatment of aortic diseases: Document covering acute and chronic aortic diseases of the thoracic and abdominal aorta of the adult. The Task Force for the Diagnosis and Treatment of Aortic Diseases of the European Society of Cardiology (ESC). Eur Heart J 2014; 35(41): 2873–926.
29. *Berretta P, Patel HJ, Gleason TG, Sundt TM, Myrmel T, Desai N,* et al. IRAD experience on surgical type A acute dissection patients: Results and predictors of mortality. Ann Cardiothorac Surg 2016; 5(4): 346–51.
30. *Rampoldi V, Trimarchi S, Eagle KA, Nienaber CA, Oh JK, Bossone E,* et al. Simple risk models to predict surgical mortality in acute type A aortic dissection: The International Registry of Acute Aortic Dissection score. Ann Thorac Surg 2007; 83(1): 55–61.
31. *Pape LA, Awais M, Woznicki EM, Suzuki T, Trimarchi S, Evangelista A,* et al. Presentation, Diagnosis, and Outcomes of Acute Aortic Dissection: 17-Year Trends From the International Registry of Acute Aortic Dissection. J Am Coll Cardiol 2015; 66(4): 350–8.
32. *Conzelmann LO, Weigang E, Mehlhorn U, Abugameh A, Hoffmann I, Blettner M,* et al. Mortality in patients with acute aortic dis-

section type A: Analysis of pre - and intraoperative risk factors from the German Registry for Acute Aortic Dissection type A (GERAADA). Eur J Cardiothorac Surg 2016; 49(2): e44–52.

33. *Osnabrugge RL, Speir AM, Head SJ, Fonner CE, Fonner E, Kappetein PA*, et al. Performance of EuroSCORE II in a large US database: Implications for transcatheter aortic valve implantation. Eur J Cardiothorac Surg 2014; 46(3): 400–8.
34. *Arnáiz-García ME, González-Santos JM, López-Rodríguez J, Dalmau-Sorlí MJ, Bueno-Codoñer M, Arévalo-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. Thorac Cardiovasc Surg 2014; 62(4): 298–306; discussion 306–7.
35. *Nezic D, Borzanovic M, Spasic T*. The external validation of the EuroSCORE II risk stratification model. Eur J Cardiothorac Surg 2014; 45: 587.
36. *Biancari F, Juvonen T, Onorati F, Faggian G, Heikkinen J, Airaksinen J,* et al. Meta-analysis on the performance of the EuroSCORE II and the Society of Thoracic Surgeons Scores in patients undergoing aortic valve replacement. J Cardiothorac Vasc Anesth 2014; 28(6): 1533–9.
37. *Chan V, Ahrari A, Ruel M, Elmistekawy E, Hynes M, Mesana TG*. Perioperative deaths after mitral valve operations may be overestimated by contemporary risk models. Ann Thorac Surg 2014; 98(2): 605–10; discussion 610.
38. *Nashef S, Sharples L*. Pride without prejudice: EuroSCORE II, the STS score and the high-risk patient subset. Eur J Cardiothorac Surg 2013; 44: 1012.
39. *Collins GS, Le Manach Y, Numata S, Itatani K, Kanda K, Yaku H*. Uninformative and misleading comparison of EuroSCORE and EuroSCORE II. Eur J Cardiothorac Surg 2017; 51(2): 399–400.
40. *Howell NJ, Head SJ, Freemantle N, Meulen TA, Senanayake E, Menon A*, et al. The new EuroSCORE II does not improve prediction of mortality in high-risk patients undergoing cardiac surgery: A collaborative analysis of two European centres. Eur J Cardiothorac Surg 2013; 44(6): 1006–11.
41. *Bai Y, Wang L, Guo Z, Chen Q, Jiang N, Dai J,* et al. Performance of EuroSCORE II and SinoSCORE in Chinese patients undergoing coronary artery bypass grafting. Interact Cardiovasc Thorac Surg 2016; 23(5): 733–9.

Received on August 10, 2017.
Revised on November 6, 2017.
Accepted on November 7, 2017.
Online First November, 2017.