



## The operative risk stratification models in cardiac surgery: EuroSCORE II model – risk groups categorization

Modeli za stratifikaciju operativnog rizika u kardiohirurgiji: kategorizacija grupa rizika za EuroSCORE II model

Duško Nežić, Miroslav Miličić, Ivana Petrović, Dragana Košević,  
Slobodan Mićović

“Dedinje” Cardiovascular Institute, Clinic of Cardiac Surgery, Belgrade, Serbia

### Abstract

**Background/Aim.** The threshold that defines a low, moderate or high-risk patients is not uniformly determined for the European System for Cardiac Operative Risk Evaluation (EuroSCORE II) by literature at present. The aim of this study was to suggest risk groups categorization within EuroSCORE II risk stratification model. **Methods.** A 7,641 consecutive patients were scored preoperatively using EuroSCORE II. The end point for the study was in-hospital mortality across the risk group categories. Patients with EuroSCORE II values of  $\leq 2.50$ ,  $> 2.50$ – $6.50\%$ , and  $> 6.50\%$  were defined to be at low, moderate, and high perioperative risk, respectively. 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 Hosmer-Lemeshow statistics, and with observed/expected (O/E) mortality ratio. **Results.** In-hospital mortality observed in our sample was 3.85% (295 out of 7,641 patients). The EuroSCORE II discriminative

power was acceptable (AUCs  $> 0.70$ ) for the low and high risk groups, while it failed to confirm good discrimination in the moderate risk group. Hosmer-Lemeshow statistics confirmed good calibration across risk group categories. The O/E mortality ratio failed to confirm good calibration in the low and high risk group (slight, but significant underprediction ratio of 1.24; 95% confidence interval 1.05–1.43), but confirmed good calibration in all three subcategories of the high risk group. **Conclusion.** The results of this study showed an acceptable overall performance of the EuroSCORE II in terms of discrimination and accuracy of model predictions for perioperative mortality across risk group categories. Validation of EuroSCORE II performances across risk group categories needs to be further studied for a continuous improvement of patients' risk stratification before planned cardiac surgery.

### Key words:

cardiac surgical procedures; hospital mortality; risk factors; risk assessment; models, theoretical; europe.

### Apstrakt

**Uvod/Cilj.** Granične vrednosti evropskog sistema za evaluaciju operativnog rizika u kardiohirurgiji (EuroSCORE II) koje definišu grupe rizika (niska, umerena, visoka) nisu trenutno uniformno određene u literaturi. Cilj naše studije bio je da se predloži kategorizacija grupa rizika u EuroSCORE II modelu. **Metode.** Preoperativno je kod 7 641 uzastopnih bolesnika procenjen preoperativni rizik upotrebom EuroSCORE II. Primarni cilj studije bio je određivanje bolesničkog mortaliteta prema pripadnosti određenoj kategoriji grupe rizika. Bolesnici sa vrednostima EuroSCORE II  $\leq 2,50$ ,  $> 2,50$ – $6,50\%$  i  $> 6,50\%$  smešteni su u kategoriju niskog, umerenog ili visokog perioperativnog rizika, respektivno. Diskriminaciona snaga modela testirana je izračunavanjem površine ispod *the receiver operating characteristic* (ROC) krive (AUC). Kalibracija modela je bila procenjena upotrebom Hosmer-Lemeshow testa, kao i sa

odnosom između zabeleženog i očekivanog (O/E) mortaliteta. **Rezultati.** Zabeleženi bolnički mortalitet u našem uzorku je bio 3.86% (295 od 7 641 bolesnika). Diskriminatorsna snaga EuroSCORE II je bila prihvatljiva za grupu niskog i visokog operativnog rizika (AUCs  $> 0,70$ ), dok dobra diskriminacija nije potvrđena u grupi umerenog rizika. Hosmer-Lemeshow testom potvrđena je dobra kalibracija u svim kategorijama grupa rizika. Odnos između zabeleženog i očekivanog (O/E) mortaliteta nije potvrdio dobru kalibraciju u grupi niskog i visokog rizika (blago, ali signifikantno potcenjivanje mortaliteta, O/E odnos od 1.24; 95% interval pouzdanosti 1,05–1,43), ali je potvrdio dobru kalibraciju u sve tri podkategorije grupe visokog operativnog rizika. **Zaključak.** Rezultati ove studije pokazuju prihvatljive sveukupne performanse EuroSCORE II modela u smislu diskriminacije i tačnosti predviđanja perioperativnog mortaliteta u definisanim kategorijama grupa rizika. Potrebna je dodatna provera performansi EuroSCORE

II u definisanim kategorijama grupa rizika zbog kontinuiranog poboljšanja stratifikacije rizika bolesnika pre planirane kardiološke intervencije.

**Ključne reči:**

hirurgija, kardijalna, procedure; mortalitet, bolnički; faktori rizika; rizik, procena; modeli, teorijski; evropa.

## Introduction

Although there has been an important progress in pre-operative screening, surgical techniques, myocardial protection, and intensive care unit (ICU) treatment, open-heart surgery still carries a certain risk of mortality and morbidity. Being the most useful tool for the improvement of patients' selection and counseling, scoring systems have been developed over the last two decades, and used to predict perioperative risk in cardiac surgery. Therefore, risk adjusted perioperative mortality rate following cardiac surgery has been widely adopted as an indicator of quality of care as well as for comparison of outcomes among institutions and surgeons (in the United Kingdom). Predicted probability of occurrence of postoperative death has also enabled stratification of patients in different clinical risk groups (low, moderate, high) <sup>1</sup>, and, subsequently, made it possible to target high-risk surgical patients in need of new therapeutic interventions <sup>2,3</sup>. Being the most widely used worldwide, the Society for Thoracic Surgeons (STS) Predicted Risk of Mortality (PROM) score, and the European System for Cardiac Operative Risk Evaluation (EuroSCORE II) have recently been adopted by guidelines <sup>4</sup>. The EuroSCORE study group <sup>5</sup> in presentation of original additive EuroSCORE model, has stratified risk groups as low (score 0–2), moderate (3–5) and high (6 ≥) perioperative risk. Although both versions (additive and logistic) of the old EuroSCORE have retained very good discriminatory power, old models no longer accurately predict operative mortality due to an overestimation of the adult cardiac patients surgical risk (poor calibration) in the range of two to three fold <sup>6,7</sup>. Therefore, the aged EuroSCORE has recently been updated and renewed into EuroSCORE II <sup>6</sup>. However, there are only a few reports <sup>8–10</sup> in which authors tried to determine risk group categories based on the score values of EuroSCORE II model. Our arbitrary determined risk group boundaries are based on predicted risk values which should represent a real world scenario, and should have a more clinically meaningful power than previously reported arithmetic quartile grouping (with similar number of patients), resulting in a very low score values for moderate, and especially for high risk patients <sup>8,9</sup>. Therefore, the aim of our study was to suggest more real risk group categorization using EuroSCORE II model.

## Methods

EuroSCORE II data were prospectively calculated (online calculator (<http://www.euroscore.org>) <sup>11</sup>, and stored in the institutional database for a series of 7,641 consecutive patients who underwent adult (≥ 18 years of age) cardiac surgery at „Dedinje“ Cardiovascular Institute in Belgrade,

Serbia, from 1st January 2012 to 31st December 2015. Due to a low number of patients with a postinfarction ventricular septal defect (VSD) included in the developmental database of EuroSCORE II, no risk coefficient was assigned to postinfarction VSD closure procedure any more <sup>6</sup>. Therefore, patients with postinfarction VSD were excluded from our study, as well as from several subsequent EuroSCORE II validation studies <sup>12,13</sup>. Only the first procedure for each patient was entered into the registry, while reinterventions for any cause in the same admission as the primary operation were coded as a complication. The primary end point for the study was in-hospital mortality (any-cause postoperative death occurring during the index hospitalization, in the hospital in which operation took the place) across the arbitrary determined risk group categories. Patients with EuroSCORE II values of ≤ 2.50, > 2.50–6.50%, and > 6.50% were defined to be at low, moderate, and high perioperative risk, respectively. High risk patients were further divided into three sub-categories – higher, very high and extremely high perioperative risk, with EuroSCORE II values of > 6.50–13.50%, > 13.50–20.00%, and > 20.00%, respectively. The Institutional Ethics Committee approved the study and requirement for informed written consent was waived due to the fact that patients' identities were masked.

Statistical analyses were performed using the statistical package SPSS version 17.0 (SPSS, Inc., Chicago, IL, USA). Categorical variables were expressed as percentages, and continuous variables were expressed as mean ± standard deviation (SD). Statistical analyses were performed by the Fisher's exact test or  $\chi^2$  test for categorical variables and by *t*-test for continuous variables. A *p*-value of less than 0.05 was considered significant.

The performance of the EuroSCORE II was analyzed focusing on discrimination power and calibration. Discrimination measures the capacity of the model to recognize the individuals of a cohort who will suffer an event (in this case perioperative death) and those who will not, thus distinguishing low-risk from high-risk patients. Discrimination can be assessed by the area under the receiver operative characteristic (ROC) 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 it is > 0.75 and good (acceptable) if it is > 0.70 <sup>14</sup>.

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 has been the most popular test to validate calibration, measuring the differences between observed and expected out-

comes over deciles of risk. A well-calibrated model gives corresponding  $p$ -value  $> 0.05$ <sup>15</sup>. We also evaluated 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 model underestimates mortality, a value below one means that 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<sup>15</sup>.

## Results

A total of 7,641 patients fulfilled the study criteria (patients  $< 18$  years of age and patients with postinfarction VSD were excluded). The baseline patients characteristics and operative details (EuroSCORE II risk factors) for our study population are presented in Table 1. There were no missing data referring to variables necessary for EuroSCORE models risk calculation. Definitions of all EuroSCORE II variables are available on the web-site: <http://www.euroscore.org><sup>11</sup>. The following subgroup procedures were performed: coronary artery bypass grafting (CABG) surgery [ $n = 4,044$  (52.9%)]; valve(s) surgery (surgery of one or more valves) [ $n = 1,732$  (22.7%)]; combined cases (CABG and valve(s) surgery) ( $n = 1,182$  (15.5%)); aortic (thoracic aorta) surgery [ $n = 545$  (7.1%)] and other major cardiac procedures [ $n = 138$  (1.8%)].

Discriminatory and calibration abilities of EuroSCORE II for the whole sample and across basic (low, moderate, high) risk group categories are summarised in Table 2.

The in-hospital mortality observed in our sample was 3.86% (295 out of 7,641 patients), while EuroSCORE II predicted mortality was 3.62%. Discriminative power of the EuroSCORE II model was determined by calculation of the AUCs. Very good discrimination was confirmed (all AUCs  $> 0.75$ ; for whole cohort and for all subgroups procedures which were performed – CABG, valve(s), combined, aortic, other). The EuroSCORE II discriminative power was acceptable (AUCs  $> 0.70$ ) for the low and high risk groups, while it failed to confirm good discrimination in the moderate risk group. In the low risk group, only subgroup of valve(s) surgery showed good discrimination, as well as two subgroups (aortic, others) in the moderate risk group. Surprisingly, almost all results in the high risk category confirmed acceptable discrimination [mostly, AUCs  $> 0.70$ ; close to borderline in the CABG subgroup (AUC = 0.69); failed in the subgroup – other]. Although H-L statistics confirmed overall good calibration in all risk group categories (overall and in all subgroups), it failed to confirm good calibration of EuroSCORE II model for the whole cohort and for subgroups of the CABG and aortic surgery (Table 2). However, the O/E mortality ratio confirmed good calibration for the whole sample, and for all subgroups of performed cardiac procedures, excluding aortic surgery (significant underestimation of mortality; O/E mortality ratio = 1.63; 95% CI 1.25–2.01). In risk group categories, the O/E mortality ratio confirmed good calibration in the moderate risk group (including all subgroups), but it failed to confirm good calibration in the low and high risk groups (whole risk group sample and aortic surgery), as well as for CABG patients in the low risk group (Table 2).

**Table 1**  
**Patients characteristics and operative details (European System for Cardiac Operative Risk II - EuroSCORE II risk factors)**

Variable	Number (%) of patients	Variable	Number (%) of patients
Age (years), mean $\pm$ SD	63.0 $\pm$ 10.5	Left ventricle function	
Gender (female)	2,269 (29.7)	good	2,890 (37.8)
Renal impairment		moderate	3,585 (45.9)
normal	3,989 (52.2)	poor	745 (9.8)
moderate	2,905 (38.0)	very poor	421 (5.5)
severe	706 (9.2)	Recent myocardial infarction	1,003 (13.1)
dialysis	41 (0.5)	Pulmonary hypertension	
Extracardiac arteriopathy	1,382 (17.4)	moderate	2,198 (28.8)
Poor mobility	54 (0.7)	severe	643 (8.4)
Previous cardiac surgery	283 (3.7)	Urgency	
Chronic lung disease	387 (5.1)	Elective	5,712 (74.8)
Active endocarditis	94 (1.2)	Urgent	1,403 (18.4)
Critical preoperative care	73 (1.0)	Emergency	518 (6.8)
Diabetics on insulin	792 (10.4)	Salvage	8 (0.1)
NYHA class		Weight of the intervention	
I	1,085 (14.2)	isolated CABG	4,044 (52.9)
II	4,303 (56.3)	single non-CABG	1,485 (19.4)
III	2,151 (28.2)	two procedures	1,502 (19.7)
IV	102 (1.3)	three procedures	610 (8.0)
CCS Class IV	481 (6.3)	surgery on thoracic aorta	545 (7.1)

NYHA – New York Heart Association; CCS – Canadian Cardiovascular Society; CABG – coronary artery bypass grafting.

**Table 2**  
**Calibration and discrimination of European System for Cardiac Operative Risk II (EuroSCORE II) across risk group categories and for the whole sample**

Risk categories – according to EuroSCORE II value	Observed mort., % (n)	Expected mort., % (n)	O / E ratio	(95% CI)	H-L <i>p</i>	AUC (95% CI)
Low (n = 4,573; 59.8%), (0.50–2.50%)						
all patients (n = 4,573)	0.92 (42)	1.32 (60)	0.70	(0.49–0.91)	0.61	0.72 (0.64–0.80)
CABG (n = 2,937)	0.65 (19)	1.28 (38)	0.50	(0.28–0.72)	0.78	0.68 (0.55–0.82)
valve(s) (n = 1,176)	0.85 (10)	1.26 (15)	0.67	(0.26–1.08)	0.35	0.75 (0.58–0.92)
combined (n = 312)	2.24 (7)	1.76 (5)	1.27	(0.33–2.44)	0.90	0.59 (0.36–0.81)
aortic (n = 53)	9.40 (5)	2.10 (1)	5.00	(1.62–9.38)	0.24	0.47 (0.22–0.72)
others (n = 95)	1.05 (1)	1.20 (1)	1.00	(-0.78–2.78)	0.40	0.39 (0.29–0.49)
Moderate (n = 2,055; 26.9%), (> 2.50–6.50%)						
all patients (n = 2,055)	4.18 (86)	3.97 (82)	1.05	(0.83–1.27)	0.76	0.62 (0.56–0.68)
CABG (n = 879)	4.89 (43)	3.81 (33)	1.30	(0.91–1.69)	0.40	0.62 (0.54–0.70)
valve(s) (n = 370)	2.97 (11)	3.87 (14)	0.79	(0.33–1.25)	0.37	0.49 (0.34–0.64)
combined (n = 510)	3.92 (20)	4.14 (21)	0.95	(0.43–1.47)	0.72	0.65 (0.52–0.78)
aortic (n = 262)	4.20 (11)	4.28 (11)	1.00	(0.41–1.59)	0.50	0.72 (0.58–0.85)
others (n = 34)	2.94 (1)	3.94 (1)	1.00	(-0.96–2.96)	0.60	0.79 (0.65–0.93)
High (n = 1,013; 13.3%), (> 6.50%)						
all patients (n = 1,013)	16.49 (167)	13.3 (135)	1.24	(1.05–1.43)	0.06	0.71 (0.67–0.75)
CABG (n = 228)	14.47 (33)	11.7 (27)	1.22	(0.80–1.64)	0.22	0.69 (0.59–0.79)
valve(s) (n = 186)	16.13 (30)	14.0 (26)	1.15	(0.74–1.56)	0.74	0.71 (0.61–0.81)
combined (n = 360)	13.30 (48)	14.1 (51)	0.94	(0.67–1.21)	0.40	0.75 (0.67–0.83)
aortic (n = 230)	23.48 (54)	13.2 (30)	1.80	(1.32–2.28)	0.11	0.72 (0.65–0.80)
others (n = 9)	22.20 (2)	13.4 (1)	2.00	(-0.77–4.77)	0.13	0.57 (0.00–1.00)
Whole sample						
all patients (n = 7,641)	3.86 (295)	3.62 (277)	1.06	(0.96–1.18)	0.001	0.84 (0.82–0.86)
CABG (n = 4,044)	2.35 (95)	2.42 (98)	0.97	(0.78–1.16)	0.001	0.84 (0.80–0.88)
valve(s) (n = 1,732)	2.94 (51)	3.19 (55)	0.93	(0.68–1.18)	0.06	0.85 (0.80–0.90)
combined (n = 1,182)	6.35 (75)	6.53 (77)	0.97	(0.86–1.08)	0.13	0.76 (0.70–0.82)
aortic (n = 545)	12.84 (70)	7.83 (43)	1.63	(1.25–2.01)	0.04	0.77 (0.71–0.84)
others (n = 138)	2.90 (4)	2.66 (4)	1.00	(0.02–1.98)	0.65	0.78 (0.00–1.00)

mort. – mortality; O – observed; E – expected; CI – confidence interval; AUC – area under curve; CABG – coronary artery bypass grafting; H-L – Hosmer-Lemeshow test.

**Table 3**  
**Calibration and discrimination of European System for Cardiac Operative Risk II (EuroSCORE II) across high risk group subcategories**

High risk subcategories according to EuroSCORE II value (n = 1013)	Observed mortality % (n)	Expected mortality % (n)	O / E ratio	(95% CI)	H-L <i>p</i>	AUC (95% CI)
Higher (n = 711; 70.2%), (>6.50–13.50%)						
all patients (n = 711)	11.25 (80)	9.11 (65)	1.23	(0.96–1.50)	0.67	0.66(0.60–0.72)
CABG (n = 182)	11.0 (20)	9.0 (16)	1.25	(0.70–1.80)	0.14	0.65(0.52–0.79)
valve(s) (n = 122)	10.7 (13)	8.95 (11)	1.18	(0.54–1.82)	0.97	0.63(0.47–0.79)
combined (n = 242)	7.85 (19)	9.3 (23)	0.83	(0.46–1.20)	0.35	0.66(0.53–0.79)
aortic (n = 160)	16.9 (27)	9.1 (15)	1.80	(1.12–2.48)	0.38	0.72(0.62–0.82)
others (n = 5)	20.0 (1)	8.4 (0.4)	N/A	N/A	0.36	0.25(0.00–1.00)
Very high (n = 167; 16.5%), (> 13.50–20.00%)						
all patients (n = 167)	19.76 (33)	16.37 (27)	1.22	(0.80–1.64)	0.01	0.52(0.41–0.63)
CABG (n = 26)	26.9 (7)	16.6 (4)	1.75	(0.45–3.05)	0.40	0.33(0.04–0.62)
valve(s) (n = 35)	17.1 (6)	16 (6)	1.0	(0.20–1.80)	0.47	0.54(0.29–0.79)
combined (n = 62)	9.7(6)	16.3 (10)	0.6	(0.12–1.08)	0.23	0.54(0.30–0.78)
aortic (n = 41)	34.1 (14)	16.7 (7)	2.0	(0.95–3.05)	0.09	0.53(0.35–0.72)
others (n = 3)	0.0 (0)	16 (0.48)	N/A	N/A	N/A	N/A
Extremely high (n = 135; 13.3%), (> 20.00%)						
all patients (n = 135)	40.0 (54)	31.7 (43)	1.24	(0.93–1.59)	0.18	0.69(0.60–0.78)
CABG (n = 20)	30.0 (6)	29.9 (6)	1.0	(0.20–1.80)	0.22	0.48(0.18–0.77)
valve(s) (n = 29)	37.9 (11)	32.8 (10)	1.10	(0.55–1.75)	0.18	0.76(0.58–0.94)
combined (n = 56)	41.1 (23)	32.1 (18)	1.28	(0.76–1.80)	0.14	0.71(0.57–0.85)
aortic (n = 29)	44.8 (13)	31.0 (9)	1.44	(1.13–1.75)	0.42	0.69(0.49–0.88)
others (n = 1)	100.0 (1)	30.6(0.3)	N/A	N/A	N/A	N/A

N/A – not applicable; O – observed; E – expected; CI – confidence interval; H-L – Hosmer-Lemeshow; CABG – coronary artery bypass grafting.

Discriminatory and calibration abilities of EuroSCORE II across high risk group subcategories are summarised in Table 3.

In the high risk group subcategories, the best discrimination was confirmed in extremely high risk group [close to borderline for the whole group and for aortic surgery (AUCs = 0.69); acceptable (AUCs > 0.70) in valves(s) and combined surgery; while it failed in CABG surgery]. In other two subcategories (higher and very high operative risk), good discrimination was recorded only for aortic surgery in higher risk group (AUC = 0.72) – Table 3.

The H-L statistics confirmed good calibration in all high risk group subcategories (higher, very high and extremely high) for all tested procedures, except for category – all patients in subcategory of very high operative risk (H-L  $p$  = 0.01) (Table 3). In the high risk group subcategories, the O/E mortality ratio failed to confirm good calibration only for aortic surgery in the higher and extremely high risk groups [O/E ratio of 1.80 (95% CI 1.12–2.48) and O/E ratio of 1.44 (95% CI 1.13–1.75), respectively] (Table 3).

## Discussion

Risk estimation is one of the most powerful tools for the improvement of the standard of care and correct allocation of clinical and economic resources<sup>16</sup>. Owing to perioperative risk stratification models, predicted probability of occurrence of perioperative death has enabled stratification of patients in different clinical risk groups (low, moderate, high), and, subsequently, made it possible to plan the optimal schedule for cardiac surgery, moderate the postoperative workload in ICU and rationally allocated hospital resources<sup>1</sup>. It has been confirmed that the additive EuroSCORE model significantly correlated with cost of cardiac surgery<sup>17</sup>, and that ICU and postoperative stay were significantly prolonged across increasing EuroSCORE II risk group categories (subsequently enhancing the cost of open heart surgery)<sup>18</sup>. Therefore, it appears that stratification in clinical risk group categories should be an integral part of the cardiac surgical practice, belonging to risk assessment, decision-making, and informed consent.

Validation of risk stratification abilities of the old, additive EuroSCORE has been conducted and presented in the basic manuscript<sup>5</sup> by the EuroSCORE study group. Validation processing confirmed good calibration for the medium risk group (score 3–5; O/E mortality ratio of 1.04; 95% CI 0.89–1.19), as well as for the high risk group (score  $\geq$  6; O/E mortality ratio of 0.99; 95% CI 0.91–1.07). For the low risk group (score 0–2; O/E mortality ratio of 0.62; 95% CI 0.42–0.82) model significantly overestimated mortality (O/E mortality ratios and 95% CIs were calculated using the data from quoted manuscript). The AUCs and H-L test  $p$ -values were not presented for risk groups.

The threshold that defines a low, moderate, high/very high-risk patients is not uniformly determined for the EuroSCORE II by literature at present. Several groups attempted to present and clarify this topic. Paparella et al.<sup>10</sup> categorised almost 6,200 patients into five risk groups (low  $\leq$  1.5%, mild

1.6–5.0%, moderate 5.1–10.0%, high 10.1–20.0%, and very high > 20%). However, that categorisation (supported by formation of a hierarchical tree, and subsequent statistical analysis) has been conducted using observed mortality, rather than predicted mortality. In our opinion, categorisation of the risk groups should be performed according to EuroSCORE II predicted mortality, and then, O/E mortality ratio and statistical analysis should be performed. Therefore, that study is not valid for EuroSCORE II risk group categorization. Velicki et al.<sup>8</sup> divided cohort of 1,247 patient in quartiles, resulting in a fact that all patients with EuroSCORE II predicted risk of more than 2.35% (4th quartile), were categorised as high-risk patients. Bai et al.<sup>9</sup> have also divided their sample (4,507 patients) in quartiles, resulting in the high-risk group (4th quartile), with EuroSCORE II value of more than 1.64%. We do believe that it is unacceptable to categorize all patients with EuroSCORE II of more than 1.64%, or even of more than 2.35%, as high-risk patients. Even with such low risk group borderlines, EuroSCORE II underestimated mortality for “high-risk group” in both papers. Two other groups reported risk group stratification, presenting EuroSCORE II values, too, but categorisation was conducted using old EuroSCORE models. Di Dedda et al.<sup>13</sup> presented a cohort of 1,090 patients, divided in quintiles of distribution, but risk stratification was created according to the old logistic EuroSCORE values. In their patient population, for the very high risk patients (observed mortality 11%), EuroSCORE II predicted mortality was 6.5% (significant underestimation). Kalender et al.<sup>19</sup> reported octagenarians (105 patients) who underwent isolated coronary artery surgery, but the old additive EuroSCORE was used for risk group categorisation. The discriminative power of EuroSCORE II model was not shown for risk group categories in any of aforementioned papers. The perioperative mortality related to cardiac surgery has decreased due to improved surgical techniques and perioperative patients management, despite sicker and more complex (baseline patients' characteristics, case mix, etc.) patients who are undergoing surgery. Although the EuroSCORE II values are generally lower (compared with additive EuroSCORE values, except for the very high risk category) for the tested group of patients<sup>18</sup>, we decided to determine borderlines for risk groups categorisation in such a way to stay close to the basic manuscript<sup>5</sup> by the EuroSCORE study group, as follows: low risk category  $\leq$  2.5% (basic manuscript 0–2%), moderate risk category > 2.5–6.5% (basic manuscript 3–5) and high risk category > 6.5% (basic manuscript  $\geq$  6). Arangalage et al.<sup>20</sup> were the only ones who searched for correspondence borderlines values for high risk patients between old logistic EuroSCORE and for EuroSCORE II, and they proposed a threshold of  $\geq$  7% of EuroSCORE II for high risk patients, which is very close to our suggested borderline value for EuroSCORE II high risk patients.

We confirmed acceptable discriminative power of EuroSCORE II in the low risk group (AUC – 0.72) and the high risk group (AUC – 0.71). In the high risk group subcategories, only for extremely high risk subcategory, discrimination was borderline acceptable (AUC – 0.69). Good discrimina-

tion was also confirmed for some subgroups of performed surgical procedures across risk group categories as well as across high risk group subcategories (Tables 2 and 3). The explanation for reduced discriminative power is statistically simple. When patients are stratified according to the risk score, and then only one strata is analyzed, the regressors and their coefficients within the stratum are different from those which allocated them to that risk group in the first place<sup>21</sup>. Therefore, we should not be surprised if discrimination drops to a lower level within the stratum<sup>21</sup>. Furthermore, a minimum of 100 (and preferably 200) events (perioperative deaths) should be included in the sample size so that model performance can be adequately assessed<sup>22</sup>. The Hosmer-Lemeshow statistics confirmed good calibration in all risk group categories and subcategories of the high risk category, and for all subgroups of performed cardiac procedures. It failed to confirm good calibration only for the whole sample (all patients) in the subcategory of very high risk patients (H-L  $p = 0.01$ ). According to O/E mortality ratio, for the low risk group model significantly overestimated mortality for the whole sample and CABG surgery subgroup, while it significantly underestimated mortality for the aortic surgery subgroup. In the moderate risk group, prediction was good for the whole sample, as well as for all subgroups of performed cardiac surgery. In the high risk group model, mortality was slightly, but significantly underpredicted for the whole sample (O/E mortality ratio – 1.24; 95% CI 1.05–1.43). On the contrary, further analysis of high risk group

subcategories confirmed good calibration for category – all patients, in all three subcategories. Therefore, our results are not in accordance with previous statements that EuroSCORE II significantly underestimates mortality in the high risk group category<sup>2, 8, 9, 13</sup>. In the high risk group category, our study is in keeping with results of Barili et al.<sup>7</sup>, who showed an optimal EuroSCORE II calibration until 30%-predicted mortality.

#### 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 7,600 patients, another limitation has been sample size, which generated relatively small specimens, including limited number of tested events (in this case perioperative deaths) for more precise subgroup analysis.

#### Conclusion

The results of this study show an acceptable overall performance of EuroSCORE II in terms of discrimination and accuracy of the model predictions for perioperative mortality across risk group categories (except overprediction of mortality in the low risk group, O/E mortality ratio). Validation of EuroSCORE II performances across risk group categories needs to be further studied for a continuous improvement of patients' risk stratification before planned cardiac surgery.

#### R E F E R E N C E S

1. Kohl P. Importance of risk stratification models in cardiac surgery. *Eur Heart J* 2006; 27(7): 68–9.
2. Howell N, Head S, Freemantle N, van der Meulen, 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; discussion 1011.
3. 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.
4. Sullivan P, Wallach J, Ioannidis J. 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(1): 1574–82.
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(1): 9–13.
6. 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.
7. 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.
8. Velicki L, Cemerlic-Adzic 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: 288–97.
9. 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: 733–9.
10. Paparella D, Guida P, Di Eusanio G, Caparrotti S, Gregorini R, Cassese M, et al. Risk stratification for in-hospital mortality after cardiac surgery: external validation of EuroSCORE II in a prospective regional registry. *Eur J Cardiothorac Surg* 2014; 46(5): 840–8.
11. *Euroscore.org (EuroScore II)*. EuroScore Study Group. 2011. Available from: <http://euroscore.org>
12. Noyez L, Kievit P, van Swieten H, de Boer M. Cardiac operative risk evaluation: the EuroSCORE II, does it make a real difference? *Neth Heart J* 2012; 20(12): 494–8.
13. Di Dedda U, Pelissero G, Agnelli B, De Vicentis C, Castelvecchio S, Ranucci M. Accuracy, calibration and clinical performances of the new EuroSCORE II risk stratification system. *Eur J Cardiothorac Surg* 2013; 43(1): 27–32.
14. Nashef S. Death and quality in cardiac surgery. *Clin Risk* 2010; 16: 130–4.
15. Nežić 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 Cardiothorac Surg* 2013; 43(1): 206.
16. Barili F, Pacini D, Rosato F, Roberto M, Battisti A, Grossi C, et al. In-hospital mortality risk assessment in elective and non-elective cardiac surgery: a comparison between EuroSCORE II and age, creatinin, ejection fraction score. *Eur J Cardiothorac Surg* 2014; 46(1): 44–8.
17. Nilsson J, Algotsson L, Hoglund P, Lubrs C, Brandt J. EuroSCORE predicts intensive care unit stay and costs of open heart surgery. *Ann Thorac Surg* 2004; 78(5): 1528–34.

18. Nežić D, Spasić T, Mirović S, Kosević D, Petrović I, Laušević-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.
19. Kalender M, Adademir T, Tasar M, Ecevit AN, Karaca OG, Salibi S, et al. Validation of EuroSCORE II risk model for coronary artery bypass surgery in high risk patients. *Kardiochir Torakochirurgia Pol* 2014; 11(3): 252–6.
20. Arangalage D, Cimadevilla C, Alkboder S, Chiampan A, Himbert D, Brochet E, et al. Agreement between the new EuroSCORE II, the logistic EuroSCORE and the Society of Thoracic Surgeons score: implications for transcatheter aortic valve implantation. *Arch Cardiovasc Dis* 2014; 107(6–7): 353–60.
21. Nashef S, Sharpley L. Pride without prejudice: EuroSCORE II, the STS score and the high-risk patient subset. *Eur J Cardiothorac Surg* 2013; 44(6): 1012.
22. Collins G, Le Manach Y. Uninformative and misleading comparison of EuroSCORE and EuroSCORE II. *Eur J Cardiothorac Surg* 2017; 51(2): 399–400.

Received on May 16, 2018.

Accepted on July 2, 2018.

Online First July, 2018.