Summary

Introduction. During the last two decades, many authors have found that European Systems for Cardiac Operative Risk Evaluation (additive and logistic models) overestimate the risk in cardiac surgery. The new European model has recently been introduced as an update to previous versions. The aim of the study was to investigate the significance of locally derived system for cardiac operative risk evaluation and to compare its predictive power with the existing European systems. Material and Methods. For developing a local risk prediction model, data from 2681 patients submitted to cardiac surgery at the Institute of Cardiovascular Diseases Vojvodina have thoroughly been collected. Logistic regression analysis was used to construct a local model for prediction of outcome. The evaluation of the local model and three European systems was performed by comparing the observed and expected hospital mortality. Results. The difference between the predicted and observed mortality regardless of the type of surgery was statistically insignificant for the additive European system (p=0.073) and the local model (p=0.134). The logistic European system overestimated the operative risk, while the new European model underestimated mortality. In coronary surgery, all models, except the logistic European system, performed well. In valvular surgery, the new European model and the local model underestimated mortality significantly, while the additive and logistic European models performed well. In combined surgery, the new European system significantly underestimated mortality (p=0.029), while the local model performed well (p=0.252). Conclusion. The locally derived model shows satisfactory results, with good calibration and discriminative power. The local model specifically outperforms all other European systems in terms of discriminatory power in combined surgery subset.

Key words: Risk Assessment; Thoracic Surgery; Mortality; Treatment Outcome

Introduction

An important component of modern cardiac surgery practice is the one of data recording, collection, and analysis for the purpose of assessing and improving the quality of service, surgical decision-making and preoperative patient education (informed consent) [1]. The outcome of a disease or surgery, in terms of survival, is obviously of great importance not only for the patient and his family but for his doctor as well [2]. Mortality is only one of the determinants of the success of an intervention. Numerous risk models for predicting postoperative mortality following a major cardiac surgery are commonly used, one of the more popular being the European System for Cardiac Operation Risk Evalu-
The EuroSCORE, in its both additive and logistic forms, has been used extensively over the last decade for the outcome prediction and hospital performance benchmarking with a new iteration of EuroSCORE having been recently presented: the EuroSCORE II [4]. Being a model developed from a large multinational cohort, the EuroSCORE II might therefore be considered a reference group incorporating different levels of outcomes. Such risk-scoring systems provide accurate prediction when the preoperative patients' characteristics and treatment profiles are comparable with those on which the system was based.

For comprehensive assessment of the role of any risk prediction model and confirmation of its applicability in contemporary cardiac surgery practice, an external validation is mandated. Moreover, the external validation is needed to assess the service provided by the specific hospital, which should be aligned with the "gold standard," that is, the EuroSCORE II. In situations where there is no acceptable alignment between the results (prediction) produced, one has the choice to: 1) continue to use one of the standard risk prediction models being aware of all the constraints; 2) adjust (recalibrate) the model; 3) develop a local model for accurate risk prediction.

In our previous paper, we advocated the development of self-made model for a number of reasons [2]. A self-made model can usually handle input data (specific patient profile, constraints and advantages of healthcare environment) more reliably yielding better risk estimation.

Since the patient profile, as well as the quality of service provided, can significantly differ among institutions and geographical areas, we sought to develop a local outcome prediction model for cardiac surgery given all the specifics of the local population as well as customized healthcare system. Our hypothesis was that a locally derived model would provide a higher level of discrimination and would be better calibrated for the Serbian population than the models derived from other populations.

The aim of the study was to investigate the significance of locally derived system for cardiac operative risk evaluation (the VojvodinaSCORE) and to compare its predictive power with the existing European systems.

Material and Methods

For the purpose of developing a local risk prediction model, data from 2681 patients submitted to cardiac surgery at the Institute of cardiovascular diseases Vojvodina (during the period from July 2011 to December 2013) have been thoroughly collected. For each patient, 53 variables were recorded and then used in statistical analyses. The evaluation of the VojvodinaSCORE and three EuroSCORE iterations (the additive, logistic and EuroSCORE II) was performed by comparing the observed and expected hospital mortality. Data were collected prospectively and analyzed retrospectively. The study was approved by the Institutional Review Board.

Three separate models (one for each surgery subtype: coronary, valvular, combined surgery) were created using binary logistic regression. The models were developed on a cohort of 1792 consecutive patients, and then validated on a cohort of 889 patients. After the validation process, the models were compacted into a single unified model named the VojvodinaSCORE.

Descriptive statistical data were compared between the groups by using either the Pearson χ² test or Fisher exact test. Continuous variables were compared between the groups by using the unpaired Student t-test or the Wilcoxon rank-sum test (depending on the normality of the distribution). A p-value of less than 0.05 was considered to be significant.

The calibration of models was assessed by using the Hosmer-Lemeshow test. A well-calibrated model gives a p-value greater than 0.05. The model discrimination was tested by means of receiver operating characteristic (ROC) curves calculating the area under the curve (AUC) – an index which was used to assess how well the model could discriminate between survivors and non-survivors. The accuracy and clinical performance of the VojvodinaSCORE was tested in the patient subgroups based on the type of cardiac operation, and the subgroups based on risk categorization.

The statistical analyses were performed with Statistical Package for the Social Sciences (SPSS) version 19.0 (SPSS Inc., Chicago, Illinois, United States) and MedCalc for Windows, version 12.2.1 (MedCalc Software, Mariakerke, Belgium).

Results

The difference between the predicted and observed mortality regardless of the type of surgery was statistically insignificant for the additive EuroSCORE (p=0.073) and the VojvodinaSCORE (p=0.134) (Table 1). The logistic EuroSCORE overestimated the observed mortality, while the EuroSCORE II underestimated mortality. The highest AUC was observed for the EuroSCORE II model, with the value of AUC not being significantly dispersed between the models.

In coronary surgery, all models except the logistic EuroSCORE performed well. The difference between the predicted and observed mortality according to the logistic EuroSCORE was significant (p=0.045), while it was not significant concerning other three models (Table 1). Again, the EuroSCORE
II yielded highest AUC (0.827) with AUC being lower than 0.8 for other three models.

In valvular surgery, the EuroSCORE II and the VojvodinaSCORE significantly underestimated mortality, while the additive and logistic EuroSCORE performed well (Table 1). AUC for the VojvodinaSCORE was around 0.7, which is substantially lower when compared to other three models (around 0.8)

In combined surgery, the difference between the predicted and observed mortality according to the additive EuroSCORE, logistic EuroSCORE and VojvodinaSCORE was not significant, while the EuroSCORE II underestimated mortality significantly (p=0.029). Discriminative power was satisfactory only for the VojvodinaSCORE (AUC – 0.752).

**Discussion**

Preoperative risk prediction models have a critical role in current cardiac surgical practice and the use of risk models to risk-stratify patients appropriately is well established [5, 6]. Choosing the most reliable model among many other models raises a question about how good the model really is in terms of effectiveness in relation to other models. In our previous studies [7–10], we analyzed the predictive value of the EuroSCORE model in coronary surgery, as well as the trends of risk factors included in the EuroSCORE model. It was observed that the profile of coronary patients undergoing surgery in one of the cardiac surgery centers is drastically changing primarily due to the significantly advanced percutaneous techniques for myocardial revascularization. In spite of the differences in the patients’ characteristics between our databases and that of the EuroSCORE, the latter performed quite well in our databases.

The EuroSCORE, in its both additive and logistic form, has been extensively used over the last decade for the outcome prediction and hospital performance benchmarking. It is generally believed

### Table 1. Cardiac operative risk evaluation using four different models

<table>
<thead>
<tr>
<th>Cardiac Surgery</th>
<th>Model Model</th>
<th>Predicted mortality</th>
<th>Observed mortality</th>
<th>P-value</th>
<th>Auroc</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Ukupno</td>
<td>Additive EuroSCORE</td>
<td>4.51</td>
<td>3.3</td>
<td>0.073</td>
<td>0.744</td>
<td>76%</td>
<td>64.9%</td>
</tr>
<tr>
<td></td>
<td>Logistic EuroSCORE</td>
<td>4.78</td>
<td>3.3</td>
<td>0.034</td>
<td>0.738</td>
<td>68%</td>
<td>72.9%</td>
</tr>
<tr>
<td></td>
<td>EuroSCORE II</td>
<td>1.86</td>
<td>3.3</td>
<td>0.002</td>
<td>0.769</td>
<td>72%</td>
<td>75.4%</td>
</tr>
<tr>
<td></td>
<td>Vojvodina SCORE</td>
<td>2.48</td>
<td>3.3</td>
<td>0.134</td>
<td>0.759</td>
<td>72.4%</td>
<td>69.6%</td>
</tr>
<tr>
<td>Coronary surgery</td>
<td>Additive EuroSCORE</td>
<td>3.83</td>
<td>2.2</td>
<td>0.064</td>
<td>0.780</td>
<td>77.8%</td>
<td>62.6%</td>
</tr>
<tr>
<td></td>
<td>Logistic EuroSCORE</td>
<td>4.00</td>
<td>2.2</td>
<td>0.045</td>
<td>0.775</td>
<td>77.8%</td>
<td>74.3%</td>
</tr>
<tr>
<td></td>
<td>EuroSCORE II</td>
<td>1.62</td>
<td>2.2</td>
<td>0.282</td>
<td>0.827</td>
<td>77.8%</td>
<td>74.6%</td>
</tr>
<tr>
<td></td>
<td>Vojvodina SCORE</td>
<td>2.29</td>
<td>2.2</td>
<td>0.376</td>
<td>0.796</td>
<td>72.7%</td>
<td>72.3%</td>
</tr>
<tr>
<td>Valvular surgery</td>
<td>Additive EuroSCORE</td>
<td>5.00</td>
<td>4.1</td>
<td>0.518</td>
<td>0.784</td>
<td>75%</td>
<td>61.7%</td>
</tr>
<tr>
<td></td>
<td>Logistic EuroSCORE</td>
<td>5.23</td>
<td>4.1</td>
<td>0.427</td>
<td>0.793</td>
<td>75%</td>
<td>61.7%</td>
</tr>
<tr>
<td></td>
<td>EuroSCORE II</td>
<td>1.87</td>
<td>4.1</td>
<td>0.010</td>
<td>0.792</td>
<td>75%</td>
<td>83.8%</td>
</tr>
<tr>
<td></td>
<td>Vojvodina SCORE</td>
<td>2.19</td>
<td>4.1</td>
<td>0.042</td>
<td>0.706</td>
<td>62.5%</td>
<td>71.1%</td>
</tr>
<tr>
<td>Combined surgery</td>
<td>Additive EuroSCORE</td>
<td>5.95</td>
<td>5.3</td>
<td>0.735</td>
<td>0.554</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Logistic EuroSCORE</td>
<td>6.61</td>
<td>5.3</td>
<td>0.516</td>
<td>0.541</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>EuroSCORE II</td>
<td>2.52</td>
<td>5.3</td>
<td>0.029</td>
<td>0.610</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Vojvodina SCORE</td>
<td>3.57</td>
<td>5.3</td>
<td>0.252</td>
<td>0.752</td>
<td>75%</td>
<td>60.8%</td>
</tr>
</tbody>
</table>
that the model shows a good level of accuracy, with a C-statistic of around 0.75 to 0.80, but could use an improvement or recalibration especially for high-risk patients [11]. The internal validation of the EuroSCORE II shows an improved C-statistic compared with the previous logistic EuroSCORE model (C-index$_{\text{EuroSCORE II}} = 0.81$ vs. C-index$_{\text{EuroSCORE}} = 0.78$) and good calibration (Hosmer–Lemeshow $\chi^2 = 15.48; p = 0.0505$) [3]. Risk models are most valid in patient populations where the preoperative patients’ characteristics and treatment protocols are comparable with those of the original environments. That is why a model should not be used elsewhere as such before its validity has been tested in the local patient material [12].

The majority of the cardiac surgery centers perform less than 1500 cardiac surgical procedures annually. Surgeons in these institutions have three possibilities: 1) to apply and use some of the existing models (“ready-made” model); 2) to recalibrate the existing model by defining new coefficients for specific factors (recalibrate); 3) to develop a completely new local model based on their experience calibrated in relation to their patient population (remodel) [13]. The later solution offers the best possibility to achieve adequate accuracy and good distinctive features of the model. Certain clinical factors, which have not been taken into account within the EuroSCORE, might have significant impact on the postoperative outcome and therefore potentially important morbidity and prognostic information may be missing. Thus, a well-calibrated model specifically designed to predict the risk for patients undergoing different cardiac surgery procedures would be applicable to improve patient selection and plan the care for high-risk patients more efficiently [14].

Due to the known constraints of ready-made models and somewhat limited results in our population published in the previous papers [2, 7, 8, 10], we have decided to create a local model. This endeavor has been implemented through support of the Provincial Secretariat for Science and Technological Development of the Autonomous Province of Vojvodina (Serbia) (Grant number 114–451–2131/2011). The main goal that we wanted to achieve was to provide our patients with the most accurate information on what their operative risk was and to provide the hospital with the most accurate tool for clinical benchmarking. Our intention was also to promote the VojvodinaSCORE as a national model for outcome prediction in cardiac surgery. Unfortunately, the idea was not accepted by others [15]. In an era of transcatheter aortic heart valve replacement when indications of valve replacement are heavily based on risk scores, the importance of having a score that accurately predicts the outcome is of paramount importance. We believe that we have developed the model capable to serve this purpose (again in conjunction with the EuroSCORE II). Clinical benchmarking and comparison of the results with other hospitals around the world is extremely important and only possible through standardized models such as the EuroSCORE. However, certain risk factors, not included in the EuroSCORE, have significant impact on the postoperative outcome.

The main limitation of our model is the relatively low sample size on which it was developed and tested. Therefore, other (external) validation would be warranted. The model reflects the experience of a single healthcare institution and may not represent national and international practice and outcomes, which may lead to a potential bias and results requiring further examination with a large number of patients across the multicenter database.

**Conclusion**

The latest iteration of the European System for Cardiac Operation Risk Evaluation outperforms earlier iterations in terms of discriminatory power, although it universally underestimates the risk. The Vojvodina System for Cardiac Operation Risk Evaluation—a locally derived model—shows satisfactory results, with good calibration and discriminative power. The locally derived model specifically outperforms all other European Systems for Cardiac Operation Risk Evaluation in terms of discriminatory power in combined surgery subset, where the European System believes that the locally derived model would be of great use in everyday clinical practice since it faithfully illustrates the actual state of patient population of the region where it was developed.

**References**


12. Pitkanen O, Niskanen M, Rehnberg S, Hippelainen M, Hynynen M. Intra-institutional prediction of outcome after card-


Rad je primljen 30. IX 2014.
Recenziran 20. X 2014.
Prihvaćen za štampu 30. X 2014.