External validation of European system for cardiac operative risk evaluation II in a Tunisian population.

Chighaly El Hadj Sidi*, Imen Mgarrech, Amine Tarmiz, Sofiane Jerbi

Department of Cardiovascular and Thoracic Surgery, Sahloul University Hospital, Sousse, Tunisia

Abstract

Objective: The main objective of this study is to evaluate the performance of the predictive model (EuroSCORE II) on a Tunisian population in order to validate its use in our country.

Methods: This is a retrospective study of data from 418 adult patients undergoing cardiac surgery with cardiopulmonary bypass between 1st January 2015 and 31 December 2016 in the department of cardiovascular and thoracic surgery of the Sahloul University Hospital of Sousse. The EuroSCORE II is calculated using the application validated on the site www.euroscore.org. The performance of the score is evaluated by analyzing its discriminative power by constructing the ROC curve and analyzing its calibration using the Hosmer-Lemeshow statistics.

Results: The EuroSCORE II shows good discriminative power in our population with an area under the ROC curve >0.7 in all study groups (0.864 ± 0.032 for general cardiac surgery, 0.822 ± 0.061 for coronary surgery, 0.864 ± 0.052 for valvular surgery, and 0.900 ± 0.041 for urgent cardiac surgery). The model appears to be calibrated as well by obtaining p values above the statistical significance level of 0.05 (0.638 for general cardiac surgery, 0.543 for coronary surgery, 0.179 for valvular surgery, and 0.082 for urgent cardiac surgery).

Conclusion: The EuroSCORE II presents acceptable performance in our population, attested by a good discriminative power and an adequate calibration.

Keywords: Cardiac Surgery, EuroSCORE II, Surgical Mortality, Discriminative power, Calibration

Introduction

In recent years, adult cardiac surgery has experienced a significant increase in operative risk due to the recruitment of an increasingly elderly population with multiple comorbidities. In addition, she benefited from improved surgical techniques and postoperative resuscitation care [1]. Despite these technical advances and accumulated knowledge, it remains a high-risk surgery, burdened with many potentially fatal complications.

The risk scores in cardiac surgery are intended to estimate the operative mortality according to the characteristics of the patient and the modalities of the surgery. They, therefore, have an important role in estimating the benefit/risk ratio of the interventions and for informing the patient, thus guiding the therapeutic choice [2]. These scores are also useful in comparing postoperative outcomes and improving the quality of care in cardiovascular facilities [3]. They have the advantage of reducing the subjectivity of the estimation of the operative risk, but must be interpreted with caution and can never be a substitute for clinical judgment.

Many predictive models have been proposed and used for cardiac surgery. The most widely used currently are the score of the STS (Society of Thoracic Surgeons), the main score applied in North America, and EuroSCORE (European System for Cardiac Operative Risk Evaluation), which is the most used model in Europe [4]. The EuroSCORE II, published in 2012, was developed on a database of 154 centers in 43 predominantly European countries [5], it must be tested and validated in developing countries such as Tunisia before being used as a model risk stratification and serve as information for the patient seeking care or as an element of monitoring and evaluation of cardiac surgery services.

To our present knowledge, EuroSCORE II has not been validated in Tunisia. In this work, we proposed to evaluate the performance of this risk stratification model (EuroSCORE II) on a Tunisian population in order to validate its use in our country.

Patients and Methods

This is an observational, transversal study conducted on a retrospective model in the Department of Cardiovascular and Thoracic Surgery at the Sahloul University Hospital of Sousse. This study focuses on adult patients undergoing cardiac surgery...
with Cardiopulmonary Bypass (CPB), over a period of 2 years from 1st January 2015 until 31 December 2016.

We included in our study all adult patients who had cardiac surgery under cardiopulmonary bypass, with or without aortic clamping. In the end, 418 patients were included in this study; they were enrolled and followed up to the 30th postoperative day.

Patient data were collected from department of archived records at Sahoul University Hospital referring to the factors in EuroSCORE II. EuroSCORE II was calculated for each patient using the validated application on www.euroscore.org.

The statistical analysis was done using SPSS software version 20.0. The quantitative data were represented as means ± standard deviations and the qualitative variables in number and percentage.

The comparison between the different data made the call to Pearson’s χ² test for proportions and Student’s T-test for averages. A univariate analysis was used to identify independent predictors of hospital mortality and a value of ρ less than 0.05 was set as the statistical significance level.

The basic overall performance parameter was the Standardized Mortality Ratio (SMR) calculated according to the formula:

\[ \text{SMR} = \frac{\text{Observed mortality}}{\text{Expected mortality}} \]

The analysis of the validity of the score was carried out by two approaches:

- Study of discriminative power by constructing the ROC curve, which has for abscissa the rate of false positive represented by the value (1-specificity) and for ordinate the rate of true positives represented by the value of the sensitivity. Thus, the area under the ROC curve (AUC) was obtained according to the method of Hanley and McNeal
- Study of the calibration using the Hosmer-Lemeshow goodness-of-fit test, and then building the calibration plot.

**Results**

**Patients’ characteristics**

The study included 418 patients underwent cardiac surgery, 245 men (58.6%) and 173 women (41.4%), with a sex ratio of 1.4. The mean age is 55.84 ± 13.84 years with extremes ranging from 18 to 87 years. Women are younger (55 ± 14 years) than men (57 ± 13 years), with no significant difference (ρ=0.09).

**Table 1. Frequencies of cardiac surgical procedures and their mortalities in our population.**

<table>
<thead>
<tr>
<th>Type of surgery</th>
<th>Frequency</th>
<th>Mortality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coronary surgery</td>
<td>160 (38.3%)</td>
<td>11 (6.8%)</td>
</tr>
<tr>
<td>Valvular surgery</td>
<td>204 (48.8%)</td>
<td>17 (8.3%)</td>
</tr>
<tr>
<td>Mixed valvulo-coronary surgery</td>
<td>16 (3.8%)</td>
<td>3 (18.7%)</td>
</tr>
<tr>
<td>Surgery of the thoracic aorta</td>
<td>26 (6.2%)</td>
<td>7 (36.7%)</td>
</tr>
<tr>
<td>Correction of congenital heart disease</td>
<td>7 (1.7%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>Resection of a heart tumor</td>
<td>4 (1%)</td>
<td>1 (25%)</td>
</tr>
<tr>
<td>Removal of an endocavity PM probe</td>
<td>1 (0.2%)</td>
<td>0 (0%)</td>
</tr>
</tbody>
</table>

These patients have undergone different types of heart surgery. Table 1 shows the frequency of various cardiac surgical procedures and their corresponding mortalities in the validation study. Table 2 how’s the distribution of risk factors in our population and their relationship to mortality (a ρ value <0.05 is considered statistically significant).

**Mortality analysis**

Of the 418 patients in our study, 39 died with a global mortality rate of 9.3%.

The mean age of the deceased patients was 60 ± 14 years versus 55 ± 14 years in the survivors without significant difference (ρ=0.361). 61.5% of them are males, while 38.5% are females without significant difference (p=0.697). The observed mortality is 6.8% in the coronary subgroup, 8.3% in the valvular subgroup and 23.3% in the urgency subgroup.

**Validation of EuroSCORE II**

**Standardized mortality ratio**

The mortality predicted by EuroSCORE II in the total population (3.25%) is significantly lower (ρ<0.001) than the observed mortality (9.3%) so that the SMR is 2.86.
In the coronary subgroup, the mortality predicted by EuroSCORE II (2.32%) is lower than the observed mortality (6.8%) without statistical significance (p=0.052), so that the SMR is 2.93. whereas in the valvular subgroup this predicted mortality (3.39%) is significantly (p<0.001) lower than the observed mortality (8.3%) with an SMR of 2.44.

The mortality predicted in the urgency sub-group (6.99%) is lower than the observed mortality (23.3%), but in a non-significant way (p=0.335) the SMR is 3.33.

The discriminative power of EuroSCORE II was estimated by the area under the ROC curve (AUC). It seems to have good discrimination in the total population as well as in all subgroups studied, the area under the curve is 0.864 ± 0.032 (between 0.801 and 0.927 with CI=95% and p<0.001) for the total population, 0.822 ± 0.061 (between 0.703 and 0.941 with CI=95% and p<0.001) for the coronary subgroup, 0.864 ± 0.052 (between 0.762 and 0.967 with CI=95% and p<0.001) for the valvular subgroup, and 0.900 ± 0.041 (between 0.819...
Figure 1. Shows the ROC curves for the total population and for the different subgroups studied.

Calibration analysis

To this extent, we performed a Hosmer-Lemeshow goodness-of-fit test that gives a $\chi^2$ value of 4.28, with a df of 6 and a $p$ value of 0.638 in the total population. This test gives a $\chi^2$ value of 2.14 with a df of 3 and a $p$ value of 0.543 for the coronary subgroup, a $\chi^2$ value of 4.90 with a df of 3 and a $p$ value of 0.179 for the valvular subgroup, and a $\chi^2$ value of 6.70 with a df of 3 and a $p$ value of 0.082 for the urgency subgroup.

The EuroSCORE II also seems to have a good calibration in the total population and the coronary subgroup, but less good in the other two subgroups ($p$ value remains greater than 0.05.

Tables 3 correspond to the contingency table of the Hosmer-Lemeshow goodness-of-fit test in the total population, while Figure 2 illustrates the calibration plots in the total population and the coronary, valvular, and urgency subgroups.

Discussion

Table 4 presents the distribution of the variables designed for the development of EuroSCORE II in our study as well as in the initial study conducted by Nashef et al.

The average age of our population was 55.84 years, while women accounted for 41.4%, against an average of 64.6 years and a female representing nearly a third of the population of EuroSCORE II [5].
Figure 2. Calibration plots.

Table 4. Comparison of patients’ characteristics between the original EuroSCORE II population and our population.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Our population (n=418)</th>
<th>Nashef [5] (n=22381)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>55.84 (± 13.84)</td>
<td>64.6 (± 12.5)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Female gender</td>
<td>173 (41.4%)</td>
<td>6919 (30.9%)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Diabetes on insulin</td>
<td>54 (12.9%)</td>
<td>1705 (7.6%)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Extracardiac arteriopathy</td>
<td>48 (11.5%)</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>Previous cardiac Surgery</td>
<td>20 (4.8%)</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>Poor mobility</td>
<td>10 (2.4%)</td>
<td>713 (3.2%)</td>
<td>=0.359</td>
</tr>
<tr>
<td>Chronic lung disease</td>
<td>12 (2.9%)</td>
<td>2384 (10.7%)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Recent MI</td>
<td>38 (9.1%)</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>Angina at rest</td>
<td>50 (12%)</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>NYHA</td>
<td>Grade II 212 (50.7%)</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Grade III 168 (40.2%)</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Grade IV 18 (4.3%)</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>Creatinine clearance</td>
<td>51-85 ml / min 156 (37.3%)</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td></td>
<td>&lt;51 ml / min 54 (12.9%)</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>Dialysis</td>
<td>4 (1%)</td>
<td>244 (1.1%)</td>
<td>=0.795</td>
</tr>
<tr>
<td>Critical preoperative state</td>
<td>12 (2.9%)</td>
<td>924 (4.1%)</td>
<td>=0.199</td>
</tr>
<tr>
<td>Ejection fraction</td>
<td>31-50% 84 (20.1%)</td>
<td>NA</td>
<td></td>
</tr>
</tbody>
</table>

These differences may be related to longer life expectancy and a lower incidence of rheumatic heart disease (more common among women in our country) than in European countries. The external validation studies of EuroSCORE II in different Western European countries [6-8] and Eastern Europe [3,9,10] gave results similar to those of the initial study, as well as than that conducted by Borracci et al. [11].

To obtain results comparable to those of our study, we can consult the Asian series like that of Atashi et al. [12], that of kar et al. [13], or that of Pillai et al. [14].

Since the population of EuroSCORE II is 9 years older than our population, there should be more comorbidities, but paradoxically, we found that diabetes on insulin was more common in our population while chronic lung disease was more common in the population of the original study. Better management of diabetes mellitus and a spread of smoking (the leading risk factor for lung diseases) in developed countries can explain these facts.

Although early surgery improves the vital prognosis in the active phase of infective endocarditis as indicated by the work of Nagai et al. [15], we found that the presence of this pathology is an independent factor of mortality after cardiovascular surgery.

A higher percentage of patients with active infectious endocarditis in our population, compared to the population of EuroSCORE II, can also be explained by the prevalence of valvular and infectious diseases in our country.

This factor has undoubtedly contributed to increasing our mortality.

With regard to emergency cardiac surgery, we found no significant difference between our results and those provided by the original study, with the exception of the rescue category, which is more common in our study. We can assume that this difference can seriously affect our results while taking into account that the notion of urgency is subjective and not yet codified.

Table 5. Review of the main results of the EuroSCORE II validation studies compared with our results.

<table>
<thead>
<tr>
<th>Author</th>
<th>Country</th>
<th>Procedure</th>
<th>n</th>
<th>SMR</th>
<th>ROC AUC</th>
<th>CI</th>
<th>χ²</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Our study</td>
<td>Tunisia</td>
<td>All</td>
<td>418</td>
<td>2.86</td>
<td>0.864</td>
<td>0.801-0.927</td>
<td>4.28</td>
<td>0.638</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Coronary</td>
<td>160</td>
<td>2.93</td>
<td>0.822</td>
<td>0.703-0.941</td>
<td>2.14</td>
<td>0.543</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Valvular</td>
<td>204</td>
<td>2.44</td>
<td>0.864</td>
<td>0.762-0.967</td>
<td>4.90</td>
<td>0.179</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Urgency</td>
<td>103</td>
<td>3.33</td>
<td>0.900</td>
<td>0.819-0.981</td>
<td>6.70</td>
<td>0.082</td>
</tr>
<tr>
<td>Nashef [5]</td>
<td>international</td>
<td>All</td>
<td>22381</td>
<td>1.05</td>
<td>0.809</td>
<td>0.782-0.836</td>
<td>15.48</td>
<td>0.0505</td>
</tr>
<tr>
<td>Koszta [3]</td>
<td>Hungary</td>
<td>All</td>
<td>2287</td>
<td>1.19</td>
<td>0.817</td>
<td>0.776-0.856</td>
<td>23.70</td>
<td>0.0084</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Coronary</td>
<td>1038</td>
<td>0.75</td>
<td>0.811</td>
<td>0.713-0.910</td>
<td>8.52</td>
<td>0.5789</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Urgency</td>
<td>593</td>
<td>1.28</td>
<td>0.791</td>
<td>0.737-0.844</td>
<td>14.18</td>
<td>0.0145</td>
</tr>
<tr>
<td>Garcia-Valentin [4]</td>
<td>Spain</td>
<td>All</td>
<td>4034</td>
<td>NA</td>
<td>0.79</td>
<td>0.76-0.82</td>
<td>38.98</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Carnero-Alcázar [7]</td>
<td>Spain</td>
<td>All</td>
<td>3798</td>
<td>1.27</td>
<td>0.851</td>
<td>0.827-0.874</td>
<td>86.69</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>
Table 5 summarizes the results of the EuroSCORE II performance analysis (SMR, discriminative power and calibration) in our study compared to those provided by the literature. By analyzing the overall performance of the model in our population, we found a generally high SMR (2.86 for the total population, 2.93 for the coronary subgroup, 2.44 for the valvular subgroup and 3.33 for the urgency subgroup), which is very different from the results published by most cardiac surgery centers in recent years, with numbers approaching 1 indicating a good performance.

Only a few series offer results comparable to ours, such as that of Stavridis et al. [9] (2.23) or that of Kar et al. [13] (1.94) for any type of cardiac surgery included, or that of Laurent et al. [16] for aortic valve replacement or that of Kalender et al. [17] (6.77) for emergency coronary surgery or that of De Oliveira et al. [18] and that of Taamallah et al. [19] for Surgery of infectious endocarditis.

Calculating the area under the ROC curve (AUC) according to the Hanley and McNeal method finds acceptable figures (0.864 for the total population, 0.822 for the coronary subgroup, 0.864 for the valve subgroup, and 0.900 for the urgency subgroup) with confidence intervals whose lower limits always greater than 0.7, which defines the threshold for the model to be discriminating [20]. These results are comparable to the statement made by the authors of the original EuroSCORE II article, which was 0.809 (0.782-0.836) [5].

The majority of the external validation studies also showed results similar to those of the initial study, while that carried out in Egypt by Amr et al. [21] in Egypt found disappointing results with an AUC of 0.52. In light of these findings, the results of our work show that the discriminating power of EuroSCORE II is adapted to our population for all groups studied: general cardiac surgery, coronary surgery, valvular surgery, and urgent cardiac surgery.

The results obtained by the Hosmer-Lemeshow goodness-of-fit test for evaluating the calibration of EuroSCORE II in our population require careful analysis. This test is currently under discussion because of its sensitivity to the number of groups and the size of the sample [22]. It was used for this study because it was used in the internal validation of the model.

The Hosmer-Lemeshow goodness-of-fit test in our population shows small $\chi^2$ values with $p$ values always above the limit for statistical significance determination which is 0.05 and this for all the subgroups studied except for the urgency subgroup whose value is approaching 0.638 for the total population, 0.543 for the coronary subgroup, 0.179 for the valvular subgroup and 0.082 for the urgency subgroup. Therefore, there is no statistically significant difference between expected mortality and observed mortality.

These results are in contradiction with those provided in the literature, which is generally in favor of a bad calibration of EuroSCORE II in parallel with the results published in the initial article [5], which shows a $p$ value (0.0505) very close to the limit of determination of statistical significance.

Other series show disappointing results with values of $p$ less than 0.05, like that of Amr et al. [21] in Egypt and that of Wang et al. [23] in China, both of which are studied in patients undergoing valvular surgery, or as the multicentric study by Grant et al. [24] made on the largest number of patients undergoing cardiac surgery in emergency (3342). These authors conclude that there is a significant difference between their observed mortality and their expected mortality.

In general, we can say that the EuroSCORE II shows a good calibration in our population subject to the small sample size.

**Conclusion**

Despite the differences in the profile of risk factors between the Tunisian population and the population constituting the database used for the development of EuroSCORE II, we can say that this risk model presents acceptable performances in our population, as evidenced by adequate discrimination and calibration.

However, we reproach him with an underestimation of the mortality especially in the patients supposed to be low risk.
At the end of this work, we proclaim the need to start prospective and especially multicentric studies on larger samples before concluding definitively on the performance of this model in our country, or even to develop an adapted version.

References


*Corresponding author:

Chighaly El Hadj Sidi
Department of Cardiovascular and Thoracic Surgery
Sahloul University Hospital, Sousse
Tunisia
E-mail: chicev27@yahoo.fr