Age, Creatinine and Ejection Fraction Score in Brazil: Comparison with InsCor and the EuroSCORE

Omar Asdrúbal Vilca Mejía, Bruna La Regina Matrangolo, David Provenzale Titinger, Leandro Batisti de Faria, Luís Roberto Palma Dallan, Filomena Regina Barbosa Calas, Luiz Augusto Ferreira Lisboa, Luís Alberto Oliveira Dallan, Fabio Biscegli Jatene

Instituto do Coração - Hospital das Clínicas - Faculdade de Medicina - Universidade de São Paulo (USP), São Paulo, SP – Brazil

Abstract

Background: Risk scores for cardiac surgery cannot continue to be neglected.

Objective: To assess the performance of “Age, Creatinine and Ejection Fraction Score” (ACEF Score) to predict mortality in patients submitted to elective coronary artery bypass graft and/or heart valve surgery, and to compare it to other scores.

Methods: A prospective cohort study was carried out with the database of a Brazilian tertiary care center. A total of 2,565 patients submitted to elective surgeries between May 2007 and July 2009 were assessed. For a more detailed analysis, the ACEF Score performance was compared to the InsCor’s and EuroSCORE’s performance through correlation, calibration and discrimination tests.

Results: Patients were stratified into mild, moderate and severe for all models. Calibration was inadequate for ACEF Score (p = 0.046) and adequate for InsCor (p = 0.460) and EuroSCORE (p = 0.750). As for discrimination, the area under the ROC curve was questionable for the ACEF Score (0.625) and adequate for InsCor (0.744) and EuroSCORE (0.763).

Conclusion: Although simple to use and practical, the ACEF Score, unlike InsCor and EuroSCORE, was not accurate for predicting mortality in patients submitted to elective coronary artery bypass graft and/or heart valve surgery in a Brazilian tertiary care center. (Arq Bras Cardiol. 2015; [online].ahead print, PP .0-0)

Keywords: Cardiac Surgical Procedures / mortality; Myocardial Revascularization; Probability; Heart Valve Diseases / surgery; Cohort Studies.

Introduction

Cardiac surgery represents a big impact on the health system due to its significant use of human and financial resources. Therefore, risk stratification becomes increasingly important.

Regarding clinical practice, there are three ways to use a risk score in a specific population. The simplest, but also the less optimal, is the immediate use of an external score, without any model adaptation. Recalibration, which maintains the same model variables, has its weights adjusted based on its own data. Remodeling is the choice of new variables based on local risk factors. Undoubtedly, this last option offers the best accuracy and the best performance.

In Brazil, the InsCor model with ten variables, a product of the remodeling of two international models, provides information on the impact of local risk factors. However, the EuroSCORE, with 17 variables, remains the most widely used in the country. Over time, controversies related to the overestimation of the EuroSCORE led to the development of the EuroSCORE II. This model, which is even more complex, had problems related to its external validation, including in Brazil, which led us to rethink about the choice of international scores and the preference for increasingly simple models.

The “Age, Creatinine and Ejection Fraction” (ACEF) score was proposed in 2009 to predict mortality in adult patients submitted to elective heart surgery. The main characteristic of this score is the fact that it is a practical one, as the ratio between age and ejection fraction is the basis of calculation, added by an additional point when preoperative creatinine is > 2.0 mg/dL. The ACEF Score was developed and validated in Italy only, where it attained good accuracy, excellent calibration and parsimony in clinical application. This model, which attained a similar performance to that of EuroSCORE, was never validated outside Italy or compared to a Brazilian model.
The objective of this study was to evaluate the performance of ACEF Score and compare its performance with the InsCor and the EuroSCORE for predicting mortality in patients undergoing coronary artery bypass graft (CABG) and/or elective heart valve surgery in a Brazilian tertiary care center.

Methods

Sample

A retrospective, observational study was carried out based on a prospective database of Instituto do Coração (InCor) of Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo (HC-FMUSP).

The hypothesis that the records included data for all variables of the chosen models had to be confirmed before the start of the analysis. Thus, to carry out the statistical validation, the sample size should include at least a hundred deaths. Data published by InCor showed a mortality of 4.8% for elective CABG and 8.4% for elective heart valve surgery; thereby, the minimum number of patients would be 2,084. We then chose the number of consecutive patients consecutively submitted to elective surgeries between May 2007 and July 2009, which amounted to a total of 2,565 patients.

Inclusion Criteria

The inclusion criteria were the following: age ≥ 18 years old and having undergone heart valve surgery (replacement or plastic), Coronary Artery Bypass Graft (CABG; with or without the use of cardiopulmonary bypass) or associated surgery (CABG and valve surgery), electively, in the established period.

Data collection, definition and organization

The collection of data from patients included in this registry was performed by a graduate student supervised by two assistant physicians of the Department of Cardiovascular Surgery of InCor. At the time, a spreadsheet was created to cover all the variables described by the 2000 Bernstein-Parsonnet and EuroSCORE models. In this analysis, 60 preoperative variables were collected per patient and placed in an interface created in Excel. After the registry evaluation, we observed that it could also provide information to validate the ACEF Score — and, moreover, it could compare its performance to that of the InsCor and EuroSCORE.

The ACEF Score value was calculated by the following equation, available in the model publication:

\[
\text{Age (y)}/\text{EF} (%) + 1 \text{ (if preoperative serum creatinine value > 2.0 mg/dL)}
\]

The additive value of InsCor was calculated from a graph with its own scoring system and the EuroSCORE, from the site: http://www.euroscore.org/calc.html. All settings attributed to the model variables (Chart 1) were taken into account, together with their respective values and classified according to their importance regarding the death event. The follow-up was limited to the hospital phase, having hospital mortality as the primary outcome, which covered the period between surgery and hospital discharge. For optimization and completeness of the data, all information recorded in the Department of Cardiovascular Surgery were confined to the Incor Data Processing Service.

Statistical analysis

To evaluate the performance of the ACEF score, the EuroSCORE and InsCor in mortality prediction, the models predictive validation was carried out in 2,565 patients. The validation was performed through calibration and discrimination tests. The calibration was calculated using the Hosmer-Lemeshow test (H-L), which shows the degree to which the model fits the data. Therefore, it has a different interpretation; if the H-L test is not significant (p > 0.05), the model has a proper fit. Models with good fit show no significance at the test, indicating that the model’s prediction is not significantly different from the observed values.

The discrimination, which differentiates high-risk patients from low-risk ones was measured by the area under the Receiver Operating Characteristics (ROC) curve. Continuous variables were expressed as mean ± standard deviation and categorical variables as percentages. The models’ performance was measured by comparing the observed and expected mortality in the risk groups established by the models. Fisher’s exact test was used for contingency tables. A p value < 0.05 was considered significant. Statistical analysis was performed with the Statistical Package for Social Sciences (SPSS) software, Statistics Desktop, version 22.0 for Windows (IBM Corporation Armonk, New York, United States).

Ethics and Informed Consent

The Free and Informed Consent was waived, as the study includes only data without identification.

Results

Sample

Of the 2,565 analyzed patients, 167 patients (6.5%) died. The mean age was 59.62 ± 13.35 years, and 913 (35.59%) patients were females. The mean ejection fraction was 56.62% ± 12.99%. The mean score values in the sample was 1.18 ± 0.54, 3.48 ± 3.31 and 3.78 ± 2.87 for the ACEF Score, InsCor and EuroSCORE, respectively. The following surgeries were performed: 1,130 (44.1%) isolated CABG; 679 (26.5%) mitral valve surgeries; 580 (22.6%) aortic valve surgeries and 176 (6.86%) CABG + valve surgeries. Nevertheless, there were 449 (17.5%) reoperation cases and 100 (3.9%) patients with creatinine > 2 mg/dL.

Adequacy of the models

There was a high positive correlation between EuroSCORE and InsCor (Spearman, r = 0.770; p < 0.0001), a moderate one between EuroSCORE and ACEF Score (Spearman, r = 0.527; p < 0.0001) and a low one between InsCor and ACEF Score (Spearman, r = 0.301; p < 0.0001) to predict mortality in the assessed sample.
The ACEF score showed association with death \( (p < 0.0001) \). However, the H-L test did not show a good model fit \( (p = 0.046) \) (Table 1). For a better analysis, the ACEF score was divided into three categories (Table 2). On the other hand, the InsCor, in addition to demonstrating an association with death \( (p < 0.0001) \), showed good fit \( (p = 0.460) \) in the H-L test (Table 3). In Table 4, the InsCor was divided into three categories. Still, the EuroSCORE showed association with death \( (p <0.0001) \) and a good fit \( (p = 0.750) \) in the H-L test (Table 5). In Table 6, the EuroSCORE was divided into three categories.

## Chart 1 - Models and their variables

<table>
<thead>
<tr>
<th>ACEF Score</th>
<th>InsCor</th>
<th>EuroSCORE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, ejection fraction and creatinine = 2 mg/dL</td>
<td>Age &gt; 70 years, female gender, associated coronary surgery + valve, recent infarction, reoperation, aortic valve surgery, tricuspid valve surgery, creatinine &gt; 2mg/dL, ejection fraction &lt; 30%, preoperative events (use of preoperative inotropics, cardiogenic shock, cardiac resuscitation, use of intra-aortic balloon, acute renal failure, heart massage, orotracheal intubation, tachycardia or ventricular fibrillation)</td>
<td>Age, gender, COPD, peripheral vascular disease, neurological dysfunction, creatinine, endocarditis, previous cardiac surgery, critical preoperative status (use of preoperative inotropics, cardiogenic shock, cardiac resuscitation and use of intra-aortic balloon), unstable angina, recent myocardial infarction, left ventricular ejection fraction, pulmonary hypertension ( (sPAP &gt; 60 \text{mmHg}) ), emergency, associated surgery, surgery on the aorta and VSD post-infarction</td>
</tr>
</tbody>
</table>

**Calibration**

The ACEF score showed association with death \( (p < 0.0001) \). However, the H-L test did not show a good model fit \( (p = 0.046) \) (Table 1). For a better analysis, the ACEF score was divided into three categories (Table 2). On the other hand, the InsCor, in addition to demonstrating an association with death \( (p < 0.0001) \), showed good fit \( (p = 0.460) \) in the H-L test (Table 3). In Table 4, the InsCor was divided into three categories. Still, the EuroSCORE showed association with death \( (p <0.0001) \) and a good fit \( (p = 0.750) \) in the H-L test (Table 5). In Table 6, the EuroSCORE was divided into three categories.

**Discrimination**

The area under the ROC curve was 0.625 (95% CI: from 0.58 to 0.67; \( p < 0.0001 \)) for the ACEF score, of 0.744 (95% CI: 0.70-0.79; \( p < 0.0001 \)) for InsCor and 0.763 (95%CI: 0.72-0.80; \( p < 0.0001 \)) for the EuroSCORE (Figure 1). There was an overlap between the ROC curve of the EuroSCORE and that of InsCor. However, the ACEF score showed the lowest area and it did not overlap with that of the other models.

**Table 1 - Association between the ACEF Score, death and the Hosmer-Lemeshow test**

<table>
<thead>
<tr>
<th>Group</th>
<th>Total</th>
<th>Death</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Yes</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Observed</td>
<td>Expected</td>
<td>Observed</td>
<td>Expected</td>
</tr>
<tr>
<td>1</td>
<td>256</td>
<td>14</td>
<td>10.26</td>
<td>242</td>
<td>245.74</td>
</tr>
<tr>
<td>2</td>
<td>256</td>
<td>9</td>
<td>11.81</td>
<td>247</td>
<td>244.19</td>
</tr>
<tr>
<td>3</td>
<td>257</td>
<td>11</td>
<td>12.72</td>
<td>246</td>
<td>244.28</td>
</tr>
<tr>
<td>4</td>
<td>257</td>
<td>7</td>
<td>13.38</td>
<td>250</td>
<td>243.62</td>
</tr>
<tr>
<td>5</td>
<td>259</td>
<td>11</td>
<td>14.20</td>
<td>248</td>
<td>244.80</td>
</tr>
<tr>
<td>6</td>
<td>255</td>
<td>21</td>
<td>14.73</td>
<td>234</td>
<td>240.27</td>
</tr>
<tr>
<td>7</td>
<td>257</td>
<td>10</td>
<td>15.76</td>
<td>247</td>
<td>241.24</td>
</tr>
<tr>
<td>8</td>
<td>257</td>
<td>25</td>
<td>17.39</td>
<td>232</td>
<td>239.61</td>
</tr>
<tr>
<td>9</td>
<td>257</td>
<td>25</td>
<td>21.35</td>
<td>232</td>
<td>235.65</td>
</tr>
<tr>
<td>10</td>
<td>254</td>
<td>34</td>
<td>35.42</td>
<td>220</td>
<td>218.58</td>
</tr>
</tbody>
</table>

**Discussion**

Risk scores should be linear and simplified formulas for predicting mortality or morbidity at the bedside and without the need for calculators or other personal assistants\(^{15}\). These instruments are usually evaluated in terms of accuracy (discriminative power), calibration and clinical performance. Many authors have emphasized that the level of precision rarely exceeds an area under the ROC curve of 0.75 (value that is inappropriate for clinical purposes), and that calibration may be inaccurate for low and high risk groups\(^{16}\). Risk scores are developed through statistical methods and subsequently validated in other populations to assess its clinical applicability. They are developed to be used in large populations, such as the EuroSCORE, constructed from 19,030 patients with nearly 450 events. This evidently allows the inclusion of a large number of independent predictors into the model \( (ten \text{events} \text{per predictor})^{17} \). However, the publication of the ACEF Score confirms the experimental hypothesis that a risk score can be developed based on a very limited number of risk factors. The evidence that the risk of mortality in elective cardiac surgery can be predicted with only three risk factors seems
to be in conflict with the general hypothesis that the higher the number of considered risk factors, the more accurate and better calibrated the model will be. However, a number of clinical, practical and mathematical considerations may explain this apparent paradox.

Wells et al\textsuperscript{18} concluded that “less is more” in the multivariate analysis. Thus, instead of including many variables, the models could be more consistent and effective when confined to a few variables (good predictors). In this study, however, the three risk factors used in the ACEF Score were also included in the InsCor and EuroSCORE. Perhaps one advantage of the ACEF Score, in addition to its simplicity, is that the variables age and ejection fraction are used in their continuous form instead of categorical, as in the other two models. The inclusion of a large number of independent variables increases the risk of multicollinearity, which, in practice, would be the risk of redundant information in the model\textsuperscript{19}. However, for individual cases, and especially in complex ones, greatly simplifying a model can be risky.

The question would be: national or international score; simple, intermediate or complex one - which one will be easier, cheaper and more appropriate?

After the publication of the administrative database of the Brazilian Unified Health System [Sistema Único de Saúde (SUS)], showing a mortality in cardiac surgery of 8\% in Brazil\textsuperscript{20}, efforts were made related to the validation of international risk models\textsuperscript{21} and the formulation of national models\textsuperscript{22}.

Comparing raw results of an administrative database with a clinical database, and also a voluntary one, such as the Society of Thoracic Surgery (STS), in the United States, is unacceptable. It should be emphasized that the STS registry.
Table 5 – Association between the EuroSCORE, death and the Hosmer-Lemeshow test

<table>
<thead>
<tr>
<th>Group</th>
<th>Total</th>
<th>Observed</th>
<th>Expected</th>
<th>Observed</th>
<th>Expected</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>294</td>
<td>5</td>
<td>3.83</td>
<td>289</td>
<td>290.17</td>
</tr>
<tr>
<td>2</td>
<td>377</td>
<td>10</td>
<td>6.79</td>
<td>367</td>
<td>370.21</td>
</tr>
<tr>
<td>3</td>
<td>280</td>
<td>6</td>
<td>6.06</td>
<td>274</td>
<td>273.04</td>
</tr>
<tr>
<td>4</td>
<td>383</td>
<td>11</td>
<td>13.11</td>
<td>372</td>
<td>369.89</td>
</tr>
<tr>
<td>5</td>
<td>332</td>
<td>15</td>
<td>15.60</td>
<td>317</td>
<td>316.40</td>
</tr>
<tr>
<td>6</td>
<td>228</td>
<td>11</td>
<td>14.63</td>
<td>217</td>
<td>213.37</td>
</tr>
<tr>
<td>7</td>
<td>232</td>
<td>18</td>
<td>20.19</td>
<td>214</td>
<td>211.81</td>
</tr>
<tr>
<td>8</td>
<td>268</td>
<td>40</td>
<td>35.73</td>
<td>228</td>
<td>232.27</td>
</tr>
<tr>
<td>9</td>
<td>171</td>
<td>51</td>
<td>50.18</td>
<td>120</td>
<td>120.82</td>
</tr>
</tbody>
</table>

Table 6 – EuroSCORE divided into mild, moderate and severe risk

<table>
<thead>
<tr>
<th>Risk</th>
<th>n</th>
<th>Observed percentage</th>
<th>95%CI</th>
<th>Expected percentage</th>
<th>LL</th>
<th>Expected percentage</th>
<th>LL</th>
<th>95%CI</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>LL</td>
<td>UL</td>
<td></td>
<td></td>
<td>LL</td>
<td>UL</td>
<td></td>
</tr>
<tr>
<td>Mild</td>
<td>951</td>
<td>2.2</td>
<td>1.3</td>
<td>3.1</td>
<td>1.9</td>
<td>1.0</td>
<td></td>
<td>2.8</td>
<td>0.750</td>
</tr>
<tr>
<td>Moderate</td>
<td>943</td>
<td>3.9</td>
<td>2.7</td>
<td>5.2</td>
<td>4.6</td>
<td>3.2</td>
<td></td>
<td>5.9</td>
<td>0.750</td>
</tr>
<tr>
<td>Severe</td>
<td>671</td>
<td>16.2</td>
<td>13.5</td>
<td>19.0</td>
<td>15.8</td>
<td>13.0</td>
<td></td>
<td>18.6</td>
<td></td>
</tr>
</tbody>
</table>

95% CI: 95% confidence interval; LL: lower limit; UL: upper limit.

Figure 1 – ROC curve for the ACEF score, InsCor and EuroSCORE in assessing the power of discrimination performed in 2565 patients. EURO: EuroSCORE; ACEF: Score: Age, Creatinine and Ejection Fraction Score.
represents 10% of the total surgeries performed annually in referral hospitals of the United States23.

In this context, the data from the public health service are important. In the capital, Rome, CAGB mortality in the same period during which the Brazilian analysis was carried out was 5.4%. When mortality was divided by socioeconomic status (at the same hospitals), mortality in the higher socioeconomic group (economically more favored and more educated) was 4.8%. In the most disadvantaged group (the poorest and least educated), mortality was 8.2%.24.

There is no doubt that SUS includes the range of poorer patients undergoing cardiac surgery in Brazil, where less than 30% of procedures are performed by the private sector25. The current scenario is extremely timely, as it provides space for the search of risk scores that are more accurate for our reality, knowing that the simpler and more effective a model is, the better cost-effectiveness of the system.

For all these reasons, the ACEF Score, validated in Italy, was chosen to be validated in Brazil, following all the recommendations of the authors on label. However, although the ACEF Score has shown to be comparable to the EuroSCORE, it showed to inferior to the InsCor and EuroSCORE and moreover, it was inadequate for our reality.

On the other hand, the EuroSCORE confirmed its good performance in Brazil26. However, as it is more complex and has the same accuracy of the InsCor (areas of overlapping ROC curves), it should be preferred exclusively for individual evaluation.

The EuroSCORE was recently remodeled9 and is now called “EuroSCORE II”. However, this version was not chosen for this study because, in addition to being more complex, it has brought to Brazil10 the same difficulties it showed in several centers worldwide26-28.

In this study, we observed that the ACEF score, even applied on label, showed its deficiency for a population with hypothetically similar characteristics to those of the population that originated it.

Thus, the accuracy of InsCor of ten variables was better than that of the ACEF score of three variables, and similar to the EuroSCORE of 17 variables, showing that intermediate complexity models may be preferred.

Limitations
First, as there was no record of patients, the calculation of ACEF score was carried out retrospectively. However, those in charge of the calculations were not aware of the outcomes. Second, although the registry includes patients from different regions of Brazil, data were collected from a single tertiary center, characterizing it as a single-center study.

Conclusion
The ACEF Score, although simple and practical, failed to predict mortality in patients submitted to coronary artery bypass graft and/or elective heart valve surgery in a Brazilian tertiary care center.

Author contributions
Conception and design of the research: Matrangolo BLR, Mejía OAV, Titinger DP, Faria LB, Dallan LRP, Lisboa LAF, Dallan LAO, Jatene FB; Acquisition of data: Matrangolo BLR, Mejía OAV, Titinger DP, Faria LB, Dallan LRP, Lisboa LAF, Dallan LAO; Analysis and interpretation of the data: Mejía OAV, Galas FRB, Jatene FB; Statistical analysis: Mejía OAV, Jatene FB; Writing of the manuscript: Mejía OAV, Faria LB, Dallan LRP, Galas FRB, Lisboa LAF, Dallan LAO, Jatene FB; Critical revision of the manuscript for intellectual content: Mejía OAV, Galas FRB, Lisboa LAF, Dallan LAO, Jatene FB.

Potential Conflict of Interest
No potential conflict of interest relevant to this article was reported.

Sources of Funding
There were no external funding sources for this study.

Study Association
This study is not associated with any thesis or dissertation work.

References


