The mortality risk stratification model including age, creatinine, and left ventricular ejection fraction (ACEF) was first described a few years ago to simplify mortality risk prediction in elective cardiac surgery\(^1\). Different investigations assessing the ACEF score performance demonstrated a discriminatory power ranging between 0.63 and 0.81 of the area under the Receiver Operating Characteristic (ROC) curve, according to center and type of surgery\(^2,5\). Further research showed that the ACEF model had lower performance than the EuroSCORE II in non-elective surgery\(^6\). Improvements in ACEF score accuracy were also obtained for specific risk groups, such as surgical ventricular reconstruction candidates\(^7\), and
a modified ACEF score incorporating creatinine clearance has been recently assessed in patients undergoing percutaneous coronary interventions, but not in cardiac surgery. Furthermore, most reports came from the same population in which the risk score was developed and validated, while validation in a different external population was rarely performed. The aim of the present study was to evaluate the efficacy of the original ACEF score and the modified ACEF model incorporating creatinine clearance to predict immediate operative mortality risk of patients undergoing elective cardiac surgery in a local population of Buenos Aires.

Materials and methods

A retrospective analysis was performed of prospectively collected data (ambispective design) over a 4-year period (2012-2015) from a consecutive series of 1190 adult patients who underwent elective cardiac surgery at the Buenos Aires University Hospital and its associated clinics. This series represented the social and demographic structure of Buenos Aires urban and suburban districts where patients came from.

All types of on-pump and off-pump elective cardiac surgery were included, except procedures associated to active valve endocarditis, post-infarction ventricular septal defect or free wall rupture, ischemic mitral regurgitation associated to acute myocardial infarction, urgent coronary surgery secondary to percutaneous coronary angioplasty complication, acute aortic dissection, trans-implantation and trans-catheter aortic valve implantation. Baseline and operative data were prospectively collected in a clinical registry, which included the variables needed to estimate the operative risk mortality based on ACEF and EuroSCORE II models, with the latter used for comparison with ACEF model performance. Additional information on other risk factors, comorbidities, major operative complications and in-hospital operative mortality was added in the computized database. Baseline preoperative clinical variables and in-hospital outcome data were recorded and analyzed with the approval of the local Institutional Review Board which waived the need for a written informed consent.

The ACEF score was calculated as originally described by Ranucci et al., using the formula age/leaf ventricular ejection fraction +1 point for serum creatinine > 2 mg/dl. The modified ACEF score was calculated with the method suggested by Garg et al., using the formula age/ejection fraction +1 point for every 10 ml/min reduction in creatinine clearance estimated with the Cockcroft-Gault (CG) equation below 60 ml/min per 1.73 m² (up to a maximum of 6 points). Therefore, a creatinine clearance between 50 to 59 ml/min per 1.73 m², 40 to 49 ml/min per 1.73 m² and 30 to 39 ml/min per 1.73 m² would receive 1, 2, and 3 additional points, respectively.

Operative mortality was defined as in-hospital mortality or mortality at 30 days after surgery for patients discharged from the hospital, including deaths occurring in rehabilitation centers, in secondary hospitals, or at home. Follow-up data after discharge were retrieved by personal or telephone contact.

Continuous variables were expressed as mean and standard deviation (SD) or as median and 25-75% percentiles (P25-P75). Kolmogorov-Smirnov (K–S) goodness-of-fit test was used to analyze normal distributions. For analysis purposes, the patient population was divided into three groups based on each score tertiles. The global accuracy of scores was evaluated in terms of calibration and discrimination. Model calibration was assessed by the Hosmer-Lemeshow goodness-of-fit test (HL2) and the Spiegelhalter Z-test. ROC curve analysis was used to estimate the performance of the ACEF and Cockcroft-Gault ACEF scores, and the EuroSCORE II to predict 30-day mortality risk. The area under the ROC curve (AUC) with its 95% confidence interval (CI) was calculated. The same analysis was repeated in the overall population and in 3 groups: isolated coronary artery bypass grafting (CABG), non-CABG surgeries, and isolated or combined aortic valve replacement (AVR). The discrimination versus calibration relationship for each score was plotted into an accuracy graph including the AUC and the HL2 or Spiegelhalter Z-tests. The cut-off values were identified at the point where the ROC curve sum of sensitivity and specificity was highest according to the Youden index. Sensitivity, specificity, positive (PPV) and negative predictive values (NPV), and overall accuracy for each risk score cut-off point were calculated. The difference between predicted and observed mortality rates were explored for different scores, using the EuroSCORE II for the risk definition in low (< 2%), intermediate (2-5%), and high (> 5%) risk groups. Operative mortality risk predicted by the ACEF model was estimated from the graphical relationship between the ACEF score and previously published expected mortality. Calibration was assessed by using the ACEF predicted mortality risk, and not by the calculated ACEF score, since results may vary significantly. Statistical comparison of observed vs. predicted mortality was done with the Z-test for standard error of the difference between means. Observed versus predicted mortality rates were compared with the χ² test. Statistical analysis was done with SPSS Statistics for Windows, Version 17.0. Chicago, SPSS, Inc. A two-tailed p value ≤ 0.05 was considered statistically significant.

Results

The population characteristics and the immediate post-operative adverse events are shown in Table 1. Overall mortality rate was 4.0% (48 observed deaths), while mean (± SD) mortality rates predicted by ACEF, ACEFCG, and EuroSCORE II models were 2.3% (± 1.8%) (p = 0.014), 6.4% (± 12.3%) (p = 0.010) and 2.5% (± 4.1%) (p = 0.038), respectively. Overall observed/predicted mortality ratio was 1.8 for the ACEF score, 0.6 for the ACEFCG score and 1.6 for EuroSCORE II. Predicted mortality based on the 3 models showed a positively skewed distribution; therefore, values were also presented as median and percentiles. In these cases, the median rates of mortality predicted by ACEF, ACEFCG, and EuroSCORE II models were 1.8% (p = 0.002), 1.9% (p = 0.003) and 1.5% (p = 0.0002), respectively. The patient population was divided into low (n = 760), intermediate (n = 326), and high (n = 104) risk groups according to the EuroSCORE II. Observed versus predicted all-cause of mortality, and observed/predicted ratio corresponding to the different risk groups of ACEF, ACEFCG, and EuroSCORE II models at 30 days are shown in Table 2. All 3 models demonstrated an adequate overall performance for the three risk groups; though the EuroSCORE II significantly underestimated mortality for the low risk group, and the ACEF score underestimated mortality for the high risk group. For the intermediate risk group, only the ACEF score significantly overestimated mortality risk.
Values of the AUC with 95% CI for each risk score and type of surgery are shown in Table 3. The ACEF<sub>CG</sub> score discriminatory power systematically improved AUC values obtained with the original ACEF score for all types of surgery; however, the EuroSCORE II showed the best AUC values. Table 4 reports the sensitivity, specificity, PPV and NPV for the cutoff values identified for each score in the overall population. The best results were observed for NPV in the 3 models. Overall accuracy was 56.1% for the ACEF score, 51.2% for the ACEF<sub>CG</sub> score and 75.9% for the EuroSCORE II.

A graphical comparison of the 3 models for each type of surgery in terms of discrimination and calibration is shown in Fig. 1. The best balances for 30-day all-cause death were provided in the following order by the EuroSCORE II and ACEF<sub>CG</sub> models in isolated or combined AVR, and in non-CABG surgeries, respectively, when assessing calibration by the HL<sup>2</sup> statistic. On the contrary, the ACEF<sub>CG</sub>
TABLE 2.– Performance of the ACEF and ACEFCG scores and the EuroSCORE II, according to risk groups

<table>
<thead>
<tr>
<th>EuroSCORE II</th>
<th>Observed mortality</th>
<th>ACEF</th>
<th>ACEFCG</th>
<th>EuroSCORE II</th>
</tr>
</thead>
<tbody>
<tr>
<td>Risk groups</td>
<td>% (95% CI)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>2.24 (1.19-3.29)</td>
<td>1.79 (1.25)</td>
<td>2.60 (0.86)</td>
<td>1.07 (2.09)</td>
</tr>
<tr>
<td>Intermediate</td>
<td>4.91 (2.56-7.25)</td>
<td>2.80 (1.75)</td>
<td>11.3 (0.43)</td>
<td>3.01 (1.63)</td>
</tr>
<tr>
<td>High</td>
<td>14.4 (7.67-21.2)</td>
<td>5.16 (2.79)</td>
<td>17.2 (0.84)</td>
<td>11.1 (1.30)</td>
</tr>
</tbody>
</table>

*Significantly underestimated (z-test)
†Significantly overestimated (z-test)

TABLE 3.– Area under the ROC curve for the three scores

<table>
<thead>
<tr>
<th>Risk score</th>
<th>Overall (n = 1190)</th>
<th>Isolated CABG (n = 640)</th>
<th>Non-CABG (n = 550)</th>
<th>AVR (n = 381)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACEF</td>
<td>0.636 (0.571-0.701)</td>
<td>0.626 (0.537-0.715)</td>
<td>0.647 (0.548-0.746)</td>
<td>0.653 (0.512-0.793)</td>
</tr>
<tr>
<td>ACEFCG</td>
<td>0.696 (0.628-0.764)</td>
<td>0.649 (0.550-0.749)</td>
<td>0.718 (0.631-0.806)</td>
<td>0.751 (0.654-0.848)</td>
</tr>
<tr>
<td>EuroSCORE II</td>
<td>0.742 (0.671-0.812)</td>
<td>0.654 (0.539-0.769)</td>
<td>0.754 (0.663-0.846)</td>
<td>0.806 (0.708-0.904)</td>
</tr>
</tbody>
</table>

TABLE 4.– Cutoff points and relative sensitivity, specificity, and positive and negative predictive values of the three scores in the overall population

<table>
<thead>
<tr>
<th>Risk score</th>
<th>Cutoff point</th>
<th>Sensitivity, %</th>
<th>Specificity, %</th>
<th>PPV, %</th>
<th>NPV, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACEF</td>
<td>1.18</td>
<td>74.5</td>
<td>56.1</td>
<td>6.6</td>
<td>98.1</td>
</tr>
<tr>
<td>ACEFCG</td>
<td>1.18</td>
<td>83.0</td>
<td>49.7</td>
<td>7.0</td>
<td>98.5</td>
</tr>
<tr>
<td>EuroSCORE II</td>
<td>2.67</td>
<td>57.4</td>
<td>76.7</td>
<td>9.2</td>
<td>97.8</td>
</tr>
</tbody>
</table>

PPV: Positive predictive value; NPV: Negative predictive value

![Fig. 1.– Predictive accuracy of the ACEF, ACEFCG and EuroSCORE II models for each group of surgeries. The position of the score in the graphs depends on the balance between the score calibration assessed by the $\chi^2$ value of the Hosmer-Lemeshow statistic (HL$\chi^2$) or the $Z_{1-0.025}$ value of the Spiegelhalter test, and the discriminatory power assessed by the area under the ROC curve (AUC). The accuracy of each score increases from top to bottom and from left to right. Circle size is proportional to the sample size of each group. The dotted line indicates a p-value threshold of 0.05 for 8 degrees-of-freedom in the $\chi^2$ distribution, or the $Z_{1-0.025}$ value, respectively for each graph.](image-url)
score had the best calibration accuracy in AVR and in CABG according to the Spiegelhalter Z-test.

Discussion

Ideally, an easier operative mortality predictive model should be advantageous in clinical surgical practice. In this sense, the joint guidelines of the European Society of Cardiology (ESC) and the European Association for Cardio-Thoracic Surgery (EACTS) have recently included the ACEF score to stratify risk in cardiac surgery. In the present study, the preoperative risk profile was assessed using the ACEF and ACEF$_{CG}$ scores, and the EuroSCORE II previously validated in Argentina. The ACEF and ACEF$_{CG}$ risk models based on only 3 predictors had worse accuracy compared with the EuroSCORE II, when applied to all types of elective cardiac surgery in our population. Nevertheless, the ACEF$_{CG}$ score systematically improved accuracy of the original ACEF score. Regarding calibration assessed by HL$_{2}^{2}$, ACEF$_{CG}$ and EuroSCORE II models had a similar good performance, though this was not confirmed by the Spiegelhalter test. The ACEF$_{CG}$ score was previously used to predict cardiac surgery outcome only in patients with left main coronary artery disease. In our study, the incorporation of creatinine clearance improved the prognostic ability of the original ACEF model for all types of cardiac surgery. On the other hand, clinical performance estimated by the ratio between observed and predicted mortality rates was fairly adequate for patients at the three risk groups, except with the EuroSCORE II model for the low risk group, the ACEF score for the high risk group, and the ACEF$_{CG}$ for the intermediate risk group. However, the reduced number of patients remaining in the intermediate and high risk groups could reduce the sample size power to detect a significant type II error. The EuroSCORE II is actually designed to predict in-hospital mortality, and the ACEF score in-hospital mortality plus mortality at 30 days after discharge. In our study, mortality was defined as in-hospital plus mortality at 30 days; therefore, there is some imbalance in the prediction analysis. Actually, the authors of the EuroSCORE II admitted that by including 30-day mortality, the death rates had to be increased by about 0.6%.$^{11}$ So, the real mortality risk prediction according to the EuroSCORE II should increase from 2.47% up to about 3%, with an overall observed/predicted mortality ratio equal to 1.31. All three scores demonstrated a very good negative predictive value (near 98%). Conversely, the positive predictive value for all three scores was very poor, not exceeding 10%. Since the models easily identified patients unlikely to die but not patients likely to die, they cannot be used to assess individual mortality risks. Notwithstanding, cutoff values for ROC curves are especially useful in transversal predictive models, usually correlating to a specific diagnosis. Consequently, its value in models discrimination assessment is limited, only showing the predicted mortality where the model is more accurate.

Only three external validations of the ACEF score have been done outside Italy. In Brazil, Mejia et al.$^{4}$ found a poor discriminatory power with an AUC of 0.63. The ACEF score performance in a French population undergoing isolated or combined aortic valve replacement showed an AUC of 0.66. Finally, Chung et al.$^{10}$ evaluated long-term outcome predictions using ACEF and ACEF$_{CG}$ scores in Asian patients with left main coronary artery disease undergoing CABG surgery. They found an AUC of 0.76 and 0.65 for all-cause death and cardiovascular mortality, respectively, with the ACEF score, whereas the accuracy with the ACEF$_{CG}$ score was 0.73 and 0.64 for each type of mortality. Although in our study, all-cause death was considered, cardiovascular origin was prevalent in presenting 30-day mortality; therefore, AUCs with ACEF models may be considered equivalent to those observed by Chung.

Notwithstanding the ACEF$_{CG}$ score showed an overall better performance than the ACEF score, the first model overestimated mortality in high-risk patients. In a recent research, Ranucci et al.$^{18}$ found that the ACEF score overestimated mortality in very high-risk cardiac surgical patients (AUC 0.52).

Recently, Chang et al.$^{19}$ found that the ACEF score satisfactorily predicted risk of acute kidney injury after mitral valve repair (AUC 0.76). They categorized renal failure severity by an increase in serum creatinine $\geq$ 0.3 mg/dl within 48 hours after surgery, or an increase in creatinine $\geq$ 1.5 times the baseline value within 7 days. In the present study, we did not categorize postoperative kidney injury by creatinine level assessment. Nevertheless, de novo dialysis in our population was three times lower than that found by Chang, though it was associated with a high mortality rate. In future research, the ACEF$_{CG}$ score should be considered a better candidate to screen patients at risk of renal failure after cardiac surgery.

A disadvantage of ACEF and ACEF$_{CG}$ scores is that predicted operative mortality risk is not directly calculated from the three variables, but must be extrapolated from an univariate association plot published by the authors.$^{1}$ The fast exponential growth of predicted mortality risk estimated from this plot may justify the overprediction of the ACEF$_{CG}$ score in higher-risk patients. Since the original chart was based on ACEF performance, a new tailored estimation for the ACEF$_{CG}$ score could be more accurate. Other limitation is that in the ACEF$_{CG}$ score, the age is included twice: age and creatinine clearance formula; hence, a co-linearity mistake may explain the miscalibration of this score.

The first limitation of the present study is its retrospective design, and that the sample represents only a portion of patients undergoing major cardiac surgery annually in Argentina. The low performance of the ACEF score in our
sample, in terms of accuracy and calibration, may be explained by the fact that internal risk models, developed and validated in the same population, may perform better than externally developed models. Differences may be also due to different proportions of low- or high-risk patients enrolled in the present series, compared with the original series of ACEF score development. Nonetheless, mortality underestimation observed in high-risk patients may be interpreted as an inadequate behavior of the model, as poor performance of surgeons, or both. Except for the ACEF score in patients undergoing CABG, all the scores demonstrated a good calibration profile assessed by the HL\(\chi^2\) test, and for every type of surgery. Nevertheless, limitations of the HL\(\chi^2\) test are well-known and include its dependence on arbitrary groupings of patients, poor power in small data sets, as well as the fact that it only results in a p-value\(^{20, 21}\). The resulting calibration patterns were different when using the HL\(\chi^2\) or Spiegelhalter Z-tests, and miscalibration was usually found with the second method for all the scores and types of surgery, except for the ACEF\(_{co}\) score in patients undergoing AVR. While it is true that the HL\(\chi^2\) test is widely accepted to assess calibration, differences found in the present study with an alternative method should be considered when evaluating score performance. Finally, significant differences between the scores, in terms of accuracy, calibration and clinical performance, may be due to the effect of a small sample size.

In conclusion, the two ACEF models demonstrated variable efficacy in predicting mortality risk, as evaluated in terms of accuracy, calibration, or clinical performance. For clinical use, the ACEF score seems to be adequate to predict mortality in low- and intermediate-risk patients. Though the ACEF\(_{co}\) score had a better discriminatory power and calibration, it tended to overestimate the expected risk. In comparison, the EuroSCORE II showed better overall accuracy than both ACEF models. Nonetheless, although the original ACEF model had a worse discriminatory capacity and calibration than the other scores, an acceptable clinical performance may be expected in predicting operative mortality in the group of low- and intermediate-risk patients, particularly in those undergoing elective non-CABG surgery and combined or isolated AVR. Since ideally, a simpler risk stratification score should be desirable for bedside clinical use, the ACEF model reasonably met the expected performance in our population.

Conflict of interests: None declared

References


---

**LA TAPA**

*Leonardo Arias, Flor 3. Técnica: óleo sobre papel entelado. Tamaño: 19 × 25 cm*


Actualmente se encuentra preparando su próxima muestra, a la cual pertenece la obra seleccionada para esta tapa.