Risk of Assessing Mortality Risk in Elective Cardiac Operations: Age, Creatinine, Ejection Fraction, and the Law of Parsimony
Marco Ranucci, Serenella Castelvecchio, Lorenzo Menicanti, Alessandro Frigiola and Gabriele Pelissero
Circulation 2009;119:3053-3061; originally published online Jun 8, 2009; DOI: 10.1161/CIRCULATIONAHA.108.842393
Circulation is published by the American Heart Association. 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2009 American Heart Association. All rights reserved. Print ISSN: 0009-7322. Online ISSN: 1524-4539

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://circ.ahajournals.org/cgi/content/full/119/24/3053

Subscriptions: Information about subscribing to Circulation is online at http://circ.ahajournals.org/subscriptions/
Permissions: Permissions & Rights Desk, Lippincott Williams & Wilkins, a division of Wolters Kluwer Health, 351 West Camden Street, Baltimore, MD 21202-2436. Phone: 410-528-4050. Fax: 410-528-8550. E-mail: journalpermissions@lww.com
Reprints: Information about reprints can be found online at http://www.lww.com/reprints
Risk of Assessing Mortality Risk in Elective Cardiac Operations
Age, Creatinine, Ejection Fraction, and the Law of Parsimony

Marco Ranucci, MD; Serenella Castelvecchio, MD; Lorenzo Menicanti, MD; Alessandro Frigiola, MD; Gabriele Pelissero, MD

Background—Several mortality risk scores exist in cardiac surgery. All include a considerable number of independent risk factors. In elective cardiac surgery patients, the operative mortality is low, the number of events recorded per year is limited, and the risk model may be overfitted. The present study aims to develop and validate an operative mortality risk score for elective patients based on a limited number of factors.

Methods and Results—The development series included 4557 adult patients who had undergone an elective cardiac operation at our institution from 2001 to 2003; the validation series includes the 4091 patients who subsequently underwent an operation. Three independent factors were included in the mortality risk model: age, creatinine, and left ventricular ejection fraction (ACEF). The ACEF score was computed as follows: age (years)/ejection fraction (%) + 1 (if serum creatinine value was >2 mg/dL). The ACEF score was compared with 5 other risk scores in the validation series. Discriminatory power (accuracy) was defined with a receiver-operating characteristics analysis. The best accuracy was achieved by the Cleveland Clinic score (0.812), with ACEF score just below it (0.808). In coronary operations, the 2 scores performed equally well (0.815 versus 0.813), and in isolated coronary operations, the best accuracy was achieved by ACEF (0.826), with the Cleveland Clinic score at 0.806.

Conclusion—A risk model limited to 3 independent predictors has similar or better accuracy and calibration compared with more complex risk scores if applied to elective cardiac operations. (Circulation. 2009;119:3053-3061.)

Key Words: risk factors ■ statistics ■ surgery

Many different mortality risk scores have been introduced and are currently in use for cardiac surgery patients. The first score, proposed by Parsonnet and colleagues1 in 1989, included 14 independent variables. Subsequently, the Cleveland Clinic score was proposed in 1992 by Higgins and colleagues2; it was dedicated to coronary artery bypass graft (CABG) procedures, with or without associated valve surgery, and included 9 independent variables. The Northern New England score3 was introduced in 1992; it was dedicated to isolated CABG procedures and based on 16 independent variables. More recently, the additive4 and the logistic5 EuroSCORE measures were introduced and subsequently widely validated.6,7 These scores are intended to be used for all adult cardiac surgical procedures, consider operative (within 30 days) mortality, and include 17 independent variables. There are also more complex risk models like the one presented by the Society of Thoracic Surgeons,8 and models dedicated to specific patient populations (CABG) covering longer periods of observation.

Clinical Perspective on p 3061

A risk score is usually evaluated in terms of accuracy (discriminatory power), calibration, and clinical performance. The above-mentioned risk scores were developed according to sound statistical methods, subsequently tested for validation in different patient populations, and found to provide an acceptable level of accuracy and clinical performance. However, many authors have highlighted that the accuracy level rarely exceeds an area under the curve (AUC) of 0.75 (a value that is inadequate for clinical purposes) and that the calibration may be poor in low- and high-risk patients.9–14 All of the risk scores mentioned above share 2 characteristics. First, they are designed to fit both elective and nonelective operations; second, they include as many factors as the underlying statistical procedure allows. These scores were developed using large patient populations; eg, the EuroSCORE was developed on 19 030 patients with a mortality rate of 4.8%, therefore accounting for ~450 events. This, of course, allows the inclusion of a very large number of independent predic-
tors in the model. However, this score is applied daily in a number of institutions performing a yearly number of operations that rarely exceeds 1500. Clinicians considering the use of a risk index within these institutions have 3 options: to simply use the existing external scores, knowing that the identified risk factors and the weight attributed to them may not correctly reflect their patient population; to adjust the weight of the risk factors on the basis of their own data; and to derive a totally new internal model from their own data and recalibrate it episodically. This last option offers the best accuracy and performance. However, if an institution is performing 1200 elective surgeries per year with a mortality rate of 3%, the yearly number of events is 36, and a model built on this basis would admit only 3 factors. It is in fact generally accepted that the number of independent variables that can be included in a multivariable logistic regression depends on the number of events, with a ratio of 10 events for each independent variable. Therefore, the update should be done every 2 or 3 years to have more events, or other (bayesian) methods should be applied to use the prior knowledge, in any case accounting for the limitation resulting from a validation data set that follows the development data set.

The experimental hypothesis of the present study is that a mortality risk score for cardiac surgery in elective procedures can be developed and may perform with adequate accuracy and calibration properties, even if the number of independent variables is limited to a maximum of 3 factors.

Methods
All data were retrieved from our institutional database after approval of the study design by the local ethics committee. Exclusion criteria were age <18 years and nonelective cardiac operation. From the original data set of 10 068 patients, 723 were excluded for being <18 years of age, and 692 were excluded for being urgent or emergency procedures. The remaining cohort of 8648 patients was included in the study. The study population comprised 2 cohorts of patients: 1 development series (4557 consecutive patients operated on in the San Donato Hospital from 2001 to 2003) and 1 validation series (4091 consecutive patients operated on in the same hospital from 2004 to 2007).

The clinical practice did not change relevantly from the development to the validation series, apart from some improvements in the cardiopulmonary bypass technique and the coagulation monitoring. Surgical teams were the same in the 2 series. The development series was analyzed with respect to the association of risk factors with operative mortality (defined as in-hospital mortality or mortality by 30 days after the operation 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 from external hospitals and rehabilitation centers or by telephone contact according to our standard practice. The follow-up was 99% completed for both the development and the validation series. The univariate association between preoperative factors (demographics, cardiovascular risk profile, laboratory assays, presence of comorbidities, operative details, and operative mortality was assessed with a relative risk analysis for categorical binary data and a Student t test for unpaired data for continuous variables. The following variables were tested for association with operative mortality: age, gender, weight, body surface area, body mass index, left ventricular ejection fraction (EF; the lowest in case of multiple recent assessments), recent (7 days) myocardial infarction, unstable angina, extracardiac arteriopathy, pulmonary hypertension, critical preoperative conditions, use of intraaortic balloon pump, active endocarditis, serum creatinine value, long-term dialytic treatment, hematocrit value, chronic obstructive pulmonary disease, medicated diabetes mellitus, neurological dysfunction, previous vascular surgery, previous cardiac surgery, operation other than isolated CABG, combined operation, and thoracic aorta operation. All of these conditions were defined, if necessary, according to the EuroSCORE definitions. Additional variables collected that are requested by other risk scores include hypertension, left ventricular aneurysm, mitral surgery, aortic surgery, 3-vessel coronary disease, left main coronary artery disease, and leukocytosis (>12 × 10^9 cells/L).

The total number of independent variables tested for association with operative mortality was 33. To avoid type I errors caused by multiple comparisons, a Bonferroni correction was applied.

All variables significant in the univariate analysis were subsequently tested for accuracy with a receiver-operating characteristics (ROC) analysis, with the AUC as a measurement of accuracy. The 3 variables with the best AUC values were used in a subsequent multivariable logistic regression analysis with Hosmer-Lemeshow statistics for calibration of the model. On the basis of the respective weights of the 3 variables (based on their regression coefficients), a mortality risk score was developed and tested for accuracy and calibration with logistic regression analysis, Hosmer-Lemeshow statistics, and ROC analysis.

Multicollinearity of the models was checked with a tolerance and inflation statistics. Linearity assumption was checked with graphical plotting of logistic log-linear regression.

In the validation series, the patient received a risk assessment using the new score developed in the first series plus 5 previously established mortality scores: additive and logistic EuroSCORE, Parsonnet, Cleveland Clinic, and Northern New England. Each patient was therefore stratified for mortality risk according to each different risk model. The same tests (logistic regression analysis, Hosmer-Lemeshow statistic, ROC analysis) were applied to establish the accuracy and calibration of each risk score. From the ROC analysis, the best cutoff values for each score were identified at the point where the sum of sensitivity and specificity was the highest according to the Youden index: (sensitivity+specificity)−1. Sensitivity, specificity, and positive and negative predictive values for each cutoff value in each risk score were calculated.

Differences between predicted and observed mortality rates were explored for different risk classes (septiles of distribution) by comparing predicted/observed event rates with 95% confidence intervals (CIs).

The same analyses were repeated in the overall population and in 2 subgroups: CABG operations with or without associated procedures and isolated CABG operations. These subgroups were selected because 2 of the 6 risk factors examined were developed on these subpopulations of patients.

---

Table 1. Development Series: Preoperative Factors Significantly Associated With Operative Mortality

<table>
<thead>
<tr>
<th>Factor</th>
<th>P</th>
<th>AUC</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>0.001</td>
<td>0.66</td>
<td>0.62–0.71</td>
</tr>
<tr>
<td>Body surface area</td>
<td>0.008</td>
<td>0.57</td>
<td>0.52–0.61</td>
</tr>
<tr>
<td>Left ventricular EF</td>
<td>0.001</td>
<td>0.57</td>
<td>0.52–0.61</td>
</tr>
<tr>
<td>Congestive heart failure</td>
<td>0.001</td>
<td>0.57</td>
<td>0.51–0.61</td>
</tr>
<tr>
<td>Serum creatinine</td>
<td>0.001</td>
<td>0.65</td>
<td>0.56–0.69</td>
</tr>
<tr>
<td>Chronic obstructive pulmonary disease</td>
<td>0.001</td>
<td>0.54</td>
<td>0.49–0.59</td>
</tr>
<tr>
<td>Hematocrit</td>
<td>0.001</td>
<td>0.62</td>
<td>0.57–0.67</td>
</tr>
<tr>
<td>Recent (7 d) myocardial infarction</td>
<td>0.019</td>
<td>0.52</td>
<td>0.48–0.58</td>
</tr>
<tr>
<td>Chronic dialysis</td>
<td>0.001</td>
<td>0.53</td>
<td>0.49–0.55</td>
</tr>
<tr>
<td>Other than isolated CABG</td>
<td>0.001</td>
<td>0.58</td>
<td>0.53–0.63</td>
</tr>
<tr>
<td>Combined operation</td>
<td>0.001</td>
<td>0.63</td>
<td>0.58–0.68</td>
</tr>
</tbody>
</table>
All tests were 2 sided. A value of $P < 0.05$ was considered significant for all statistical tests; after Bonferroni correction, this value was placed at 0.0015 for tests including multiple comparisons. Statistical calculations were performed with a computerized statistical program (SPSS 11.0, SPSS Inc, Chicago, Ill).

The authors had full access to and take full responsibility for the integrity of the data. All authors have read and agree to the manuscript as written.

**Results**

**Development of the Model**

The development series of 4557 patients was used for the following analyses. In this series, there were 150 operative deaths (3.3%; 95% CI, 2.4 to 3.8).

The univariate association between risk factors and operative mortality was assessed with a univariate analysis. Eleven variables were significantly associated with operative mortality (Table 1). For each variable, an ROC analysis was performed, with estimation of the AUC and 95% CI. The 3 risk factors with the best AUC were age, left ventricular EF, and serum creatinine value. These 3 factors were subsequently entered into a multivariable logistic model (stepwise forward) and confirmed to be independent predictors of operative mortality (Table 2). This multivariable model was tested for calibration with a Hosmer-Lemeshow $\chi^2$ and found to be well calibrated. Multicollinearity diagnostics with tolerance statistics excluded multicollinearity of the 3 factors.

The above model was subsequently translated into a mortality risk score. Mortality risk was directly correlated with age and inversely correlated with EF. These factors were checked for linearity assumption, and they both were demonstrated to be not linearly associated with mortality rate. The logit relationship between age and mortality (Figure 1A) is U shaped and follows a cubic equation, whereas the relationship between EF (expressed as percentage) and mortality is logarithmic (Figure 1B). The 2 factors were merged into a single combined factor, which is simply the ratio between age (years) and EF (%). This score has a range from $\approx 0.25$ (age

![Figure 1. Linearity assumption for the graphic assessment of the relationships between mortality and age, mortality and EF, mortality and age/EF, and mortality and ACEF.](circ.ahajournals.org)

Table 2. Development Series: Multivariable Logistic Regression Analysis for Operative Mortality With the 3 Identified Predictors

<table>
<thead>
<tr>
<th>Factor</th>
<th>$b$</th>
<th>SE</th>
<th>$P$</th>
<th>Odds Ratio</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>0.062</td>
<td>0.011</td>
<td>0.001</td>
<td>1.064</td>
<td>1.04–1.09</td>
</tr>
<tr>
<td>EF</td>
<td>-0.056</td>
<td>0.007</td>
<td>0.001</td>
<td>0.945</td>
<td>0.93–0.96</td>
</tr>
<tr>
<td>Serum creatinine</td>
<td>0.271</td>
<td>0.052</td>
<td>0.001</td>
<td>1.31</td>
<td>1.18–1.45</td>
</tr>
<tr>
<td>Constant</td>
<td>-5.240</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$b$ indicates regression coefficient. Hosmer-Lemeshow $\chi^2$, 5.60; $P = 0.692$. Multicollinearity diagnostics: age, 0.990 tolerance; EF, 0.991 tolerance; and serum creatinine, 0.988 tolerance.
of 18 years and EF of 70%) to ~4.0 (age of 80 years and EF of 20%). In a graphical analysis, the linearity assumption for this factor was confirmed (Figure 1C). The odds ratio for mortality as determined by logistic regression analysis was 3.95 (95% CI, 3.3 to 4.7) per each point of the age/EF score.

Serum creatinine value was dichotomized according to a cutoff value of 2.0 mg/dL. This value, identified by analyzing the coordinates of the ROC curve for mortality risk according to the serum creatinine value, was associated with a relative risk of mortality of 5.3 (95% CI, 3.5 to 7.8). This odds ratio is close to that identified for each point of the age/EF score. Therefore, 1 additional point was attributed to this condition (serum creatinine ≥2.0 mg/dL). The final 3-factor risk score was therefore calculated as follows: age/EF + 1 (if serum creatinine ≥2.0 mg/dL). This mortality risk score, based on age (A), creatinine (C), and EF, was defined as the ACEF score. Linearity assumption was checked and confirmed for the ACEF score (Figure 1D).

The ACEF score was tested for significance, accuracy, and calibration using a logistic regression model with Hosmer-Lemeshow $\chi^2$ and ROC analysis (Table 3). The ACEF was significantly ($P<0.001$) correlated with operative mortality and demonstrated very good calibration ($\chi^2$, 4.97; $P=0.791$). The accuracy was moderately good, with an AUC of 0.744. The $r^2$ correlation value was 0.82. The graphical relationship between ACEF score and operative mortality risk is reported in Figure 2.

### Validation of the Model

The validation series of 4091 patients was used to validate the ACEF score and compare its performance against 5 previously established mortality risk scores. There were 105 operative deaths in this series (2.6%; 95% CI, 2.1 to 3.1, not significantly different from the development series).

In this series, ACEF score was assessed, along with the Cleveland Clinic, Parsonnet, and Northern New England scores and the additive and logistic EuroSCORE. Given the elective nature of our patient population, some of the risk

---

**Table 3. Development Series: Predictive Function of ACEF Score for Operative Mortality (Logistic Regression)**

<table>
<thead>
<tr>
<th>Factor</th>
<th>$b$</th>
<th>SE</th>
<th>$P$</th>
<th>Odds Ratio</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACEF</td>
<td>1.24</td>
<td>0.10</td>
<td>0.001</td>
<td>3.45</td>
<td>2.28–4.31</td>
</tr>
<tr>
<td>Constant</td>
<td>−5.41</td>
<td>0.21</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$b$ indicates regression coefficient. Hosmer-Lemeshow $\chi^2$, 4.97; $P=0.791$; AUC at ROC analysis, 0.744; 95% CI, 0.699 to 0.789; SE: 0.023; $P<0.001$. $r^2=0.82$. 

AUC at ROC analysis, 0.744; 95% CI, 0.699 to 0.789; SE: 0.023; $P<0.001$. $r^2=0.82$.

---

**Figure 2.** Univariate association (logistic regression) between ACEF score and mortality risk.
factors included in these last 5 risk scores were obviously not present, including emergency case for the Cleveland Clinic score; failed angioplasty, preoperative use of intraaortic balloon pump, and salvage operation for the Parsonnet score; urgent or emergency surgery for the Northern New England score; and urgent surgery, critical preoperative conditions, and post–myocardial infarction ventricular septum defect repair for the additive and logistic EuroSCOREs.

The 6 mortality scores were tested for their accuracy in predicting operative mortality with ROC analysis in the overall population and separately for CABG operations (with or without associated procedures) and isolated CABG operations. Calibration was tested with a Hosmer-Lemeshow χ² test applied to the logistic regression equation linking each risk score to the operative mortality.

Figures 3 through 5 report the ROC curves of the 6 risk scores. Values of the AUC, with the relative 95% CIs, are reported in Table 4. Table 5 reports the sensitivity, specificity, and positive and negative predictive values for the cutoff values identified.

The ACEF had a very good accuracy in all subsets of the population, with values always exceeding 0.8. It was the second-best predictor for accuracy (after the Cleveland Clinic score) in the overall population and in patients receiving CABG with or without associated procedures and the best predictor (AUC, 0.826) in patients receiving an isolated CABG operation.

All the risk scores demonstrated a very good negative predictive value (99%). Conversely, the positive predictive value for all the risk scores was very poor, not exceeding 8%.

Calibration was assessed separately for the 6 scores and for the different subsets of patient population. In the overall population, calibration was poor for the logistic EuroSCORE (P=0.005) and the Northern New England score (P=0.049); in patients undergoing CABG with or without associated procedures, calibration was good for all 6 scores; and in patients undergoing isolated CABG, calibration was poor for the Northern New England score (P=0.04).

A final performance test was applied to the ACEF score and the additive/logistic EuroSCORE. Patients in the validation series were divided into ACEF septiles of distribution, each including ~580 patients, resulting in 7 mortality risk classes. Predicted and observed mortality rates for each risk class were estimated and compared (Table 6). Except for the first septile, in which ACEF score significantly overestimated mortality risk, there were no significant differences between predicted and observed mortality rates. Conversely, the logistic EuroSCORE overestimated the mortality risk in the first 6 septiles, and the additive EuroSCORE overestimated in the first 5 septiles and underestimated in the seventh septile.

Discussion

The results of this study confirm the experimental hypothesis that a mortality risk score can be developed on the basis of a very limited number of risk factors in elective cardiac operations. Using 3 risk factors, we were able to develop a risk model with an accuracy equivalent to or even better than more complex models, with good calibration and satisfying clinical performance.

The evidence that mortality risk in elective cardiac operations can be predicted with just 3 risk factors seems to conflict with the general feeling that the more risk factors we consider
and include in a risk model, the more accurate and better calibrated this model will be. However, we believe that a number of clinical, practical, and mathematical considerations may justify this apparent paradox.

The Problem of Overfitting
When numerous variables are included in an attempt to “control” or “adjust” the data, the accuracy of the results may be threatened, and the general advice of the statisticians is to be parsimonious in selecting independent variables. To avoid including too many independent variables, some authors even suggest “clustering” together variables with similar clinical meaning (eg, New York Heart Association class, symptoms of heart failure, anaerobic thresholds at the cardiopulmonary test). Wells and colleagues concluded that “...less is more in multivariable analysis. Instead of including all the many variables that might be statistically significant, the analysis can be more consistent and effective if confined to the few variables or pre-selected combinations of variables that are the most powerful predictors.” Of course, if a simple model can explain a phenomenon with the same level of accuracy as complex models, according to the “law of parsimony” or “the Ockham razor” concept, this model should be preferred, at least until a better-performing complex model appears.

The Problem of Risk Factor Definition
The 3 risk factors used in the ACEF score are continuous variables. Two of them (age and serum creatinine value) are definitely not subject to personal estimation. The third (EF) could be less defined because the patient may reach the operating room with different values reported in different examinations (angiography or echocardiography) or even within the same examination done at different times. However, differences are usually limited, and we decided to consider the most recently measured EF value or the lowest EF value in cases with multiple values available immediately before the operation. This gives a standardized assessment of this risk factor.

Moreover, the crucial importance of EF is highlighted in the ACEF, in which every possible value is included in the final score. By doing so, we can reach higher accuracy in the range of very low EF values. Nowadays, it is certainly possible to operate on patients with an EF <25% or even 20%. All other risk scores fail to discriminate between these very poor EF values, simply categorizing them in a field <30% (Parsonnet score and EuroSCORE), 35% (Cleveland Clinic score), or 40% (Northern New England score).

The ACEF score does not include any categorical binary risk factors. Conversely, a number of risk factors of this kind are included in the other risk scores, namely chronic obstructive pulmonary disease, cerebrovascular disease, extracardiac arteriopathy, and unstable angina. All of these factors require a definition to be included in the model. These definitions are clearly stated in each risk score, but they are subject to a certain degree of personal interpretation. Therefore, it is possible that different operators will provide different interpretations, resulting in a different final risk score, as has been demonstrated by other authors. This factor may introduce accuracy and calibration problems that are absent in the ACEF score.

The Problem of Multicollinearity
Multicollinearity is defined as the intercorrelation between independent variables included in the risk model. The risk factors included in the ACEF have been tested for multicol-

Table 4. Validation Series: ROC Curve Details for the 6 Risk Scores

<table>
<thead>
<tr>
<th>Risk Score</th>
<th>Overall (n=4091) AUC (95% CI)</th>
<th>CABG (n=2775) AUC (95% CI)</th>
<th>Isolated CABG (n=2256) AUC (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cleveland Clinic</td>
<td>0.812 (0.764–0.859)</td>
<td>0.815 (0.759–0.871)</td>
<td>0.806 (0.728–0.885)</td>
</tr>
<tr>
<td>ACEF</td>
<td>0.808 (0.761–0.854)</td>
<td>0.813 (0.756–0.869)</td>
<td>0.826 (0.754–0.897)</td>
</tr>
<tr>
<td>Logistic EuroSCORE</td>
<td>0.793 (0.746–0.840)</td>
<td>0.798 (0.742–0.853)</td>
<td>0.766 (0.682–0.851)</td>
</tr>
<tr>
<td>Additive EuroSCORE</td>
<td>0.787 (0.737–0.837)</td>
<td>0.788 (0.730–0.847)</td>
<td>0.745 (0.656–0.834)</td>
</tr>
<tr>
<td>Northern New England</td>
<td>0.785 (0.737–0.833)</td>
<td>0.800 (0.749–0.851)</td>
<td>0.804 (0.733–0.875)</td>
</tr>
<tr>
<td>Parsonnet</td>
<td>0.768 (0.721–0.816)</td>
<td>0.763 (0.711–0.816)</td>
<td>0.744 (0.673–0.815)</td>
</tr>
</tbody>
</table>

Table 5. Validation Series: Cutoff Values and Relative Sensitivity, Specificity, and Positive and Negative Predictive Values of the 6 Risk Scores in the Overall Population

<table>
<thead>
<tr>
<th>Risk Score</th>
<th>Cutoff Value</th>
<th>Sensitivity, %</th>
<th>Specificity, %</th>
<th>PPV, %</th>
<th>NPV, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cleveland Clinic</td>
<td>4.5</td>
<td>73</td>
<td>76</td>
<td>7.7</td>
<td>99</td>
</tr>
<tr>
<td>ACEF</td>
<td>1.54</td>
<td>71</td>
<td>78</td>
<td>8</td>
<td>99</td>
</tr>
<tr>
<td>Logistic EuroSCORE</td>
<td>6.3</td>
<td>72</td>
<td>73</td>
<td>6.6</td>
<td>99</td>
</tr>
<tr>
<td>Additive EuroSCORE</td>
<td>6.5</td>
<td>70</td>
<td>74</td>
<td>6.7</td>
<td>99</td>
</tr>
<tr>
<td>Northern New England</td>
<td>5.75</td>
<td>73</td>
<td>72</td>
<td>6.3</td>
<td>99</td>
</tr>
<tr>
<td>Parsonnet</td>
<td>11.5</td>
<td>73</td>
<td>64</td>
<td>5</td>
<td>99</td>
</tr>
</tbody>
</table>

PPV indicates positive predictive value; NPV, negative predictive value.
linearity, and no intercorrelation was found. We cannot address the point of intercorrelation in the other risk scores, but the inclusion of a large number of independent variables increases the risk of multicollinearity. From a clinical point of view, it is likely that intercorrelation exists between variables like age and anemia; obesity and diabetes mellitus; age and neurological dysfunction; recent myocardial infarction and unstable angina; urgent procedure, unstable angina, and critical preoperative state; and many others. These variables are included in other risk scores, making them more prone to the risk of multicollinearity. In practical terms, this means that there is a risk of redundant information being included in the model.

### The Problem of Assessing Risk in Both Elective and Emergent Conditions

The prerequisite of ACEF score is that it is applicable only to elective surgical procedures. Although this could be seen as a limitation, we must keep in mind that elective procedures are the great majority of cardiac operations. Within our database, they represent about 80% of the patient population. A large number of risk factors included in the other risk scores do not apply to this population (urgent or emergent procedure, critical preoperative state, preoperative use of intraaortic balloon pump, post–myocardial infarction ventricular septum defect repair, and some acute endocarditis cases). Therefore, the ACEF is probably better fitted to elective cases than the other risk scores, which actually try to provide, with a single tool, a risk estimate for very different conditions. A scoring system that is applied to the entire range of possible clinical conditions is probably offering a good “mean” prediction but inevitably suffers from poor accuracy and calibration at the extremes of the range (the very-low- and very-high-risk conditions). A number of studies addressed this point with respect to the additive and logistic EuroSCORE, highlighting the risk of overestimation of mortality risk in the overall population and in subsets of patients.9–14

### Inclusion/Exclusion of Risk Factors

We understand that a simple risk score based on age, serum creatinine, and EF excludes many possible risk factors. The exclusion of a large number of comorbidities and character-istics of the operation (CABG, mitral valve, combined procedure, ascending aorta, etc) could be a cause for concern. However, if we consider the comorbidities included by other risk scores and excluded by the ACEF, the truth is that there is not general agreement about which should be included and which excluded. Chronic obstructive pulmonary disease and diabetes mellitus are included in 3 of the 4 existing risk scores; obesity, peripheral arteriopathy, and cerebrovascular disease are included in 2 of the 4. Anemia, hypertension, prior vascular surgery, leukocytosis, and many others are included in just 1 risk score. The only risk factor that is always present is a repeat surgery. However, in our development series, this factor did not reach statistical significance in determining operative mortality and was not considered for our model.

Finally, it may be surprising that the ACEF attributes no role to the type of operation performed. One could argue that it is well known that mitral valve procedures carry a higher risk than isolated CABG and that CABG plus mitral valve procedures are associated with the worst outcome within elective operations. However, if we look again at the other existing risk scores, we must admit that the same defect could be attributed to the widely used EuroSCORE, which does not take into account mitral valve procedures and provides a specific risk score only to ascending aorta operations and post–myocardial infarction ventricular septum repair. All other operations are simply defined as “nonisolated CABG”; therefore, this risk score considers mortality risk the same for an isolated aortic valve replacement and a CABG plus mitral valve procedure.

Except in the case of risk scores dedicated to specific cardiac operations (like the Cleveland Clinic and the Northern New England scores), the problem of attributing a specific risk to different surgical procedures is far from being solved.

### Study Limitations

The present study includes only elective patients; therefore, it cannot be applied to the whole patient population. Moreover, the outcome variable (operative mortality) may not reflect the “real” mortality after cardiac operations; the risk of dying after cardiac operations actually levels after 3 months for coronary operations and 6 months for valve procedures. Finally, our results are not validated in different hospitals or with different patient populations that could include specific risk conditions based on epidemiological or geographical factors. This introduces the problem of different possible strategies when a risk model is applied to individual institutions: using an available risk score, adjusting it, or creating a totally new score that could take into consideration the existence of specific risk profiles within each patient population. Our approach is based on a logistic regression model, and we cannot exclude that other systems based on different statistical concepts (like bayesian models) may result in the same or better accuracy using the same limited number of factors.

Finally, one of the possible advantages of risk stratification is that some factors may be modified by therapeutic measures.
and consequently the risk may be reduced. This applies to some factors that are excluded by the ACEF (chronic obstructive pulmonary disease, pulmonary hypertension) in which age is a surrogate for many comorbidities except renal dysfunction. Consequently, the ACEF has a limited power for triggering preoperative therapeutic measures.

It is not the intention of this study to conclude that the existing risk scores are devoid of statistical soundness or clinical usefulness. They all have a good reputation, have been used for years, and remain important. Moreover, having been developed for mortality risk assessment in a patient population inclusive of elective, urgent, and emergent procedures, they bear the burden of assessing mortality risk in a very heterogeneous clinical setting. The ACEF has been developed to assess mortality risk in elective procedures, which has the advantage of limiting the variability of the prediction. However, elective procedures constitute the great majority of cardiac surgery operations done every day all over the world. Of course, this simple tool needs further validation studies in different hospitals and different countries.

This study follows the philosophical concept of the law of parsimony (“entities must not be multiplied beyond necessity”). This approach has strong supporters and detractors, like Neal, who conversely believes that “Sometimes a simple model will outperform a more complex model. . . . Nevertheless, I believe that deliberately limiting the complexity of the model is not fruitful when the problem is evidently complex. Instead, if a simple model is found that outperforms some particular complex model, the appropriate response is to define a different complex model that captures whatever aspect of the problem led to the simple model performing well.”

Actually, in our series, a very simplified model, the ACEF, performed equivalently to a more complex model, the EuroSCORE, that is currently considered the gold standard in many European countries. This equivalence could be seen as “both the models are performing well” or “both the models are bad performers.” Actually, the point is that all the risk scores have a very good negative predictive value but a very poor positive predictive value. This means that it is easy to identify patients unlikely to die but still very difficult to identify patients likely to die. Therefore, the existing risk scores cannot be used to assess individual mortality risk. This is probably due to the fact that some factors or extreme risk conditions are not included in many of them (morbid obesity, uncompensated diabetes mellitus, metabolic syndrome, genetic thrombophilic patterns, and many others). The fact that these factors are not included in the ACEF and in the other risk scores does not mean that they do not influence mortality risk. They may affect this risk, but because of the reduced density of the patients carrying this variability and the existing uncertainty of the alternative models, the impact of these variables on the calibration and discriminatory power is minimal.

This statement means that the present standard is unsatisfactory, leading to the consequence that different, and probably even more complex, models are required to increase the positive predictive value for operative mortality. A good starting point could be represented by other existing models, which gave good evidence of efficacy in CABG patients.

Disclosures

None.

References

Clinical Perspective

Risk stratification in cardiac surgery is a relevant issue. Many mortality risk models are currently available for clinicians and institutions. They are used to assign a specific operative mortality risk to individuals, to provide internal comparisons between surgeons, and to contrast the hospital performance with external benchmarks. Internal models accounting for specific risk profiles of the patient population may perform better than externally developed models. However, whereas external models are generally developed using very large series, often including up to 10,000 patients, internal models often rely on a more limited patient population. When addressing elective patients, the mortality rate is low, and the number of events is limited, allowing few predictors to be included in the model. In this retrospective, observational study, we have developed and validated an operative mortality risk score based on only 3 predictors: age, creatinine value, and left ventricular ejection fraction. This score provided a level of accuracy similar to or better than the EuroSCORE and other mortality risk scores with a better clinical performance. The limited number of predictors included in the model allows its use in relatively limited patient series, with the possibility of an annual recalibration, therefore accounting for possible local changes in clinical practice and risk profile of the patient population.