Performance of the European System for Cardiac Operative Risk Evaluation II: A meta-analysis of 22 studies involving 145,592 cardiac surgery procedures

Pietro Guida, PhD,a Florinda Mastro, MD,a Giuseppe Scrascia, MD, PhD,a Richard Whitlock, MD, PhD, FRCS,b,c and Domenico Paparella, MDa

Objectives: A systematic review of the European System for Cardiac Operative Risk Evaluation (euroSCORE) II performance for prediction of operative mortality after cardiac surgery has not been performed. We conducted a meta-analysis of studies based on the predictive accuracy of the euroSCORE II.

Methods: We searched the Embase and PubMed databases for all English-only articles reporting performance characteristics of the euroSCORE II. The area under the receiver operating characteristic curve, the observed/expected mortality ratio, and observed-expected mortality difference with their 95% confidence intervals were analyzed.

Results: Twenty-two articles were selected, including 145,592 procedures. Operative mortality occurred in 4293 (2.95%), whereas the expected events according to euroSCORE II were 4802 (3.30%). Meta-analysis of these studies provided an area under the receiver operating characteristic curve of 0.792 (95% confidence interval, 0.773-0.811), an estimated observed/expected ratio of 1.019 (95% confidence interval, 0.899-1.139), and observed-expected difference of 0.125 (95% confidence interval, −0.269 to 0.519). Statistical heterogeneity was detected among retrospective studies including less recent procedures. Subgroups analysis confirmed the robustness of combined estimates for isolated valve procedures and those combined with revascularization surgery. A significant overestimation of the euroSCORE II with an observed/expected ratio of 0.829 (95% confidence interval, 0.677-0.982) was observed in isolated coronary artery bypass grafting and a slight underestimation of predictions in high-risk patients (observed/expected ratio 1.253 and observed-expected difference 1.859).

Conclusions: Despite the heterogeneity, the results from this meta-analysis show a good overall performance of the euroSCORE II in terms of discrimination and accuracy of model predictions for operative mortality. Validation of the euroSCORE II in prospective populations needs to be further studied for a continuous improvement of patients’ risk stratification before cardiac surgery. (J Thorac Cardiovasc Surg 2014;148:3049-57)

Supplemental material is available online.

Predictive models are used to generate a risk of event according to the patient risk factor profile. In cardiac surgery, the focus has been to identify available and measurable risk factors that best predict the mortality. The European System for Cardiac Operative Risk Evaluation (euroSCORE) group has developed multivariable models for quantifying the impact of multiple risk factors on cardiac surgery operative mortality.1-3 The initial score system was recently updated with new data used to generate the model named “euroSCORE II.” The new system, published in February 2012, showed higher discrimination and better calibration than the previous model.5 To date, the performance of the euroSCORE II has been assessed in several studies. However, generalization of results assessing the external performance of the euroSCORE II is limited by differences between populations that may influence the predictive accuracy of a scoring system applied to local circumstances. An appropriate systematic evaluation of the euroSCORE II should include all evidence and results of available individual studies. We performed a meta-analysis (1) to assess euroSCORE II discriminatory power estimated by the area under the curve (AUC) receiver operating characteristic (ROC) and (2) to compare observed versus euroSCORE II predicted mortality.
METHODS

Literature Search
A systematic review of published evidence on the performance value of the euroSCORE II was undertaken using published guidelines for meta-analysis of observational studies. On February 26, 2014, one investigator (FM) searched Embase and PubMed sources for all articles reporting performance characteristics of the euroSCORE II. The free-text “euroSCORE II” was used as search term for finding the articles published after January 2012. Moreover, we evaluated bibliographies of retrieved studies and all articles citing the original euroSCORE II study obtained from the European Journal of Cardio-Thoracic Surgery website.

Study Selection
The inclusion criteria were as follows: full-text English-only articles evaluating predictive performance of the euroSCORE II on perioperative mortality in cardiac surgery. Two co-authors (FM and PG) independently reviewed each abstract. According to a standardized data-extraction form, they read all the abstracts to identify the potentially eligible articles and then used the full text of these articles to determine whether they could be included. Discordant opinions were resolved by discussion with the participation of a third investigator (DP). Duplicate publications were excluded. In the circumstance that multiple articles reported on the same dataset, the published article with the greatest number of patients was selected. Articles with very specific domains were excluded (ie, endocarditis or aortic dissection).

Parameters Extracted
The number of patients, their mean age, and the proportion of female gender were extracted. The following parameters were considered: mortality, the AUC of the ROC, and the euroSCORE II mean value. We attempted to obtain missing data from the authors by electronic mail to addresses cited in articles. The periods of enrollment, number of centers, types of surgeries performed, and definition of mortality were reported for each study.

Statistical Analysis
We calculated the observed mortality according to the number of events reported in the original studies. To standardize the estimation of the 95% confidence interval (CI) among studies, the mortality rate was estimated by using the normal approximation of binomial distribution and the 95% CI of AUC was estimated using an approximate formula. The mortality observed/expected (O/E) ratio and observed-expected (O-E) were calculated for each study. A value of O/E greater than 1 means that the euroSCORE II model underestimated relative mortality, whereas a value less than 1 means that the model overestimated mortality. If the 95% CI of the O-E difference excluded the value 0, it was considered statistically significant. Parameters analyzed were the AUC of the ROC, O/E ratio, and O-E difference with their 95% CI. The forest plot of AUC, O/E ratio, and O-E difference with their 95% CIs were displayed. Heterogeneity among studies was evaluated by the chi-square test ($P < .05$). Baujat plot was used as a graphic method to identify studies that affect heterogeneity or the meta-analytic estimate. We conducted a meta-regression analysis based on study-level data: Mean age, number of centers involved (multicenter vs single-center study), proportion of isolated coronary artery bypass grafting (CABG), presence of patients operated before euroSCORE II study population, and mortality definition (in-hospital vs 30 days) were considered as explanatory variables. The random-effects approach was used to account for unexplained heterogeneity between studies. To evaluate the predictive accuracy of euroSCORE II according to the number of centers involved in each study, a sensitivity of the findings was examined by separately performing the analysis with multicenter and single-center studies. Subgroups analysis was focused on isolated CABG, isolated valve surgery, and combined procedures (CABG with valve surgery). Moreover, high-risk surgery was evaluated by extracting procedures with greater preoperative score or nonelective surgery. The analyses were made using STATA software, version 12 (StataCorp LP, College Station, Tex).

RESULTS

Literature Search
Figure 1 presents the flow diagram of study selection. A total of 135 and 374 citations were retrieved from PubMed and Embase databases, respectively. Among all these articles, 31 studies were considered for more detailed evaluation. Inclusion and exclusion criteria were fulfilled by 24 studies that were selected for data extraction and analysis; 22 studies were included in the main analysis,3,7-27 and 2 studies were included only for subgroup analysis28,29 because they reported on surgery categories extracted from datasets of other included articles.10,23

Characteristics of the Studies Included in the Meta-Analysis
Table 1 shows the principal characteristics of the studies included in the meta-analysis.3,7-29 The number of participants who were assessed in the studies ranged from...
### TABLE 1. Main characteristics of the studies included in the meta-analysis

<table>
<thead>
<tr>
<th>Study (reference)</th>
<th>Enrollment (y)</th>
<th>Centers (n)</th>
<th>Patients (n)</th>
<th>Age (y)</th>
<th>Female (%)</th>
<th>Mortality (%)</th>
<th>euroSCORE II (%)</th>
<th>AUC</th>
<th>Surgery</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nashef and colleagues</td>
<td>2010</td>
<td>154</td>
<td>5553</td>
<td>64.6</td>
<td>30.9</td>
<td>4.2</td>
<td>4.0</td>
<td>0.81</td>
<td>Cardiac surgery</td>
</tr>
<tr>
<td>Biancari and colleagues</td>
<td>2006-2011</td>
<td>1</td>
<td>1027</td>
<td>67.0</td>
<td>22.2</td>
<td>2.7</td>
<td>4.5</td>
<td>0.87</td>
<td>Isolated CABG</td>
</tr>
<tr>
<td>Di Dedda and colleagues</td>
<td>2010-2011</td>
<td>1</td>
<td>1090</td>
<td>64.5</td>
<td>31.7</td>
<td>3.8</td>
<td>3.1</td>
<td>0.81</td>
<td>Cardiac surgery</td>
</tr>
<tr>
<td>Chalmers and colleagues</td>
<td>2006-2010</td>
<td>1</td>
<td>5576</td>
<td>69.3</td>
<td>28.6</td>
<td>3.4</td>
<td>4.5</td>
<td>0.79</td>
<td>Cardiac surgery</td>
</tr>
<tr>
<td>Grant and colleagues</td>
<td>2010-2011</td>
<td>41</td>
<td>23,740</td>
<td>67.1</td>
<td>27.7</td>
<td>3.1</td>
<td>3.4</td>
<td>0.81</td>
<td>Cardiac surgery</td>
</tr>
<tr>
<td>Carnero-Alcázar and colleagues</td>
<td>2005-2010</td>
<td>1</td>
<td>3798</td>
<td>67.0</td>
<td>37.7</td>
<td>5.7</td>
<td>4.5</td>
<td>0.85</td>
<td>Cardiac surgery</td>
</tr>
<tr>
<td>Kunt and colleagues</td>
<td>2004-2012</td>
<td>1</td>
<td>11,700</td>
<td>67.3</td>
<td>34.1</td>
<td>4.8</td>
<td>4.4</td>
<td>0.83</td>
<td>Cardiac surgery</td>
</tr>
<tr>
<td>Qadir and colleagues</td>
<td>2006-2011</td>
<td>41</td>
<td>12,470</td>
<td>67.9</td>
<td>18.2</td>
<td>1.9</td>
<td>3.0</td>
<td>0.79</td>
<td>Isolated CABG</td>
</tr>
<tr>
<td>Wang and colleagues</td>
<td>2010-2011</td>
<td>1</td>
<td>372</td>
<td>—</td>
<td>2.1</td>
<td>1.9</td>
<td>0.70</td>
<td>Isolated CABG</td>
<td></td>
</tr>
<tr>
<td>Grant and colleagues</td>
<td>2010-2011</td>
<td>41</td>
<td>3116</td>
<td>—</td>
<td>2.1</td>
<td>2.6</td>
<td>0.77</td>
<td>Isolated AVR</td>
<td></td>
</tr>
<tr>
<td>Grant and colleagues</td>
<td>2010-2011</td>
<td>41</td>
<td>2401</td>
<td>—</td>
<td>4.4</td>
<td>4.8</td>
<td>0.72</td>
<td>AVR and CABG</td>
<td></td>
</tr>
<tr>
<td>Osnabrugge and colleagues</td>
<td>2003-2012</td>
<td>17</td>
<td>6191</td>
<td>67.4</td>
<td>34.1</td>
<td>7.8</td>
<td>6.3</td>
<td>0.79</td>
<td>Isolated CABG</td>
</tr>
<tr>
<td>Osnabrugge and colleagues</td>
<td>2003-2012</td>
<td>7</td>
<td>1201</td>
<td>67.6</td>
<td>28.1</td>
<td>12.2</td>
<td>8.5</td>
<td>0.82</td>
<td>Nonelective surgery</td>
</tr>
</tbody>
</table>

(Continued)
216 to 50,588. Among the 145,592 patients (weighted mean age, 64.6 years; 31.2% were female), there were 4293 (2.95%) observed deaths and 4802 (3.30%) expected deaths. Fifteen studies reported in-hospital mortality as the main outcome. Of the 136,286 patients in these studies, 3871 (2.84%) died before the discharge and 4453 (3.27%) deaths were expected.

**Discrimination and Predictive Accuracy**

Figure 2 shows the forest plot of the AUC of the ROC, O/E ratio, and O-D difference for the studies included in the main analysis. Statistical heterogeneity was detected across studies. As a whole, our data demonstrate that the euroSCORE II has good overall ability to differentiate between patients who died and survived with an AUC of 0.792 (95% CI, 0.773-0.811). The accuracy of predictions was estimated with an O/E ratio of 1.019 (95% CI, 0.899-1.139) and O-D difference of 0.125 (95% CI, -0.269 to 0.519). The overall discrimination and predictive accuracy of euroSCORE II were similar between studies that evaluated hospital mortality by including only patients operated after the euroSCORE II cohort and those that evaluated 30-day mortality or included patients with a completely retrospective evaluation of the euroSCORE II: AUC 0.814 (95% CI, 0.800-0.828) versus 0.783 (95% CI, 0.759-0.807), O/E ratio 1.020 (95% CI, 0.908-1.133) versus 1.030 (95% CI, 0.878-1.181), and O-D difference 0.074 (95% CI, -0.331 to 0.479) versus 0.163 (95% CI, -0.333 to 0.659). Figure E1 shows a Baujat plot to evaluate the impact of studies on heterogeneity and on the meta-analytic estimate. As displayed in Figure E1, A, 2 studies with low AUC contributed to the variability and overall estimate, and the study with the highest discrimination affected heterogeneity. The largest study involved in this meta-analysis had an impact on heterogeneity, and another wide study deviated from the overall estimate because of underestimation of the predicted risk (Figure E1, B and C). Both studies were large retrospective evaluations of patients undergoing operation over a 10-year period. Studies that evaluated hospital mortality by including only the most recent patients operated after the euroSCORE II cohort had a low impact on both heterogeneity and overall result (Figure E1, group B).

**Sensitivity Analysis**

When the 9 multicenter studies (112,799 patients) were analyzed, the discrimination remained preserved with an AUC of 0.786 (95% CI, 0.760-0.812) without deviation between observed and expected mortality: The O/E ratio was 0.925 (95% CI, 0.797-1.053), and O-D difference was −0.215 (95% CI, −0.602 to 0.172). The analysis of the 13 single-center studies estimated an AUC of 0.801 (95% CI, 0.775-0.827), an O/E ratio of 1.146 (95% CI, 0.915-1.377), and an O-D difference of 0.517 (95% CI, −0.254 to 1.289).

**Meta-Regression Analysis**

Multiple regression analysis (Table 2) did not show a significant association between the euroSCORE II performance parameters (AUC ROC, O/E ratio, or O-D difference) and the study-specific data (mean age, multicenter vs single-center category, presence of patients operated before 2010, proportion of isolated CABB, and definition of mortality used).

**Subgroup Analysis**

Figure 3 shows the forest plot of AUC, O/E ratio, and O-D difference for isolated CABB, isolated valve surgery, and combined CABB and valve procedures. Statistical heterogeneity was detected across studies for O/E ratio and O-D difference (P < .01 in each subgroup and overall). The euroSCORE II showed overall good discrimination in isolated CABB (AUC, 0.804; 95% CI, 0.771-0.837) and valve surgery (AUC, 0.777; 95% CI, 0.744-0.811). A lower AUC was observed in combined procedures (0.735; 95% CI, 0.704-0.767). The agreement between observed and expected mortality was good for isolated valve procedures: O/E ratio 0.925 (95% CI, 0.753-1.096) and O-D difference −0.193 (95% CI, −0.681 to 0.296). No deviation between estimated mortality and predicted risk was...
observed in the combined surgery group: O/E ratio 1.030 (95% CI, 0.798-1.262) and O/E difference 0.176 (95% CI, −0.301 to 0.649). An overestimation of the euroSCORE II was observed for isolated CABG surgery: O/E ratio 0.829 (95% CI, 0.677-0.982) and O/E difference −0.397 (95% CI, −0.846 to 0.051).

High-Risk Procedures
When nonelective surgery or patients at greater preoperative risk were analyzed as high risk,12-14,17,20,23-25,27,28 the euroSCORE II had a lower discrimination estimated by an AUC of 0.752 (95% CI, 0.676-0.828) and a nonsignificant underestimation of prediction: O/E ratio 1.253 (95% CI, 0.936-1.570) and O/E difference 1.859 (95% CI, −0.615 to 4.332).

DISCUSSION
Prediction models play an important role in current cardiac surgical practice, and the euroSCORE II is one of the several available tools to assess the risk of perioperative mortality. Among the numerous models that have been developed for risk stratification, the euroSCORE II is the most recent risk assessment system developed using widely available clinical variables and operation-related factors.3 The score is an update of the old logistic euroSCORE model and is derived from a more contemporary data set that better reflects current cardiac surgical practice. We undertook a meta-analysis of euroSCORE II overall performance by assessing the available literature and combining results from different populations to obtain summary estimates with greater precision and statistical power than individual studies.

There has been recent interest in developing a more robust and objective scoring system to cover a wider population of cardiac surgical patients.30-32 The most frequently used risk calculators are the Society of Thoracic Surgeons (STS) Risk Score in North America and the euroSCORE in Europe. The STS Risk Score was most updated in 2008 from a large multicenter database in the United States. The STS Risk Tool was derived separately for CABG, valve procedures, and combined CABG with valve surgery to predict mortality and morbidity.30-32 In comparison with the STS Risk Score, the euroSCORE II was developed over a lower number of patients involved...
in the derivation cohort and included less variables and interaction terms between risk factors for the prediction of early mortality after different types of cardiac surgery.\textsuperscript{3,30-32} The euroSCORE II was derived from a database of more than 20,000 consecutive patients from 154 hospitals in 43 countries who underwent operation from May to July 2010. Information on risk factors was collected and compared by patients’ outcomes in terms of hospital mortality. By means of logistic regression analysis, those risk factors that were robust in predicting mortality became part of the euroSCORE II calculation.\textsuperscript{3} Several authors have evaluated the external validity of the euroSCORE II to test the generalizability of this prediction model.\textsuperscript{7-29,33,34} We conducted a meta-analysis to evaluate the discrimination of predictions and the agreement with mortality. By combining results from different studies, our meta-analysis of these studies shows that the euroSCORE II had good ability in distinguishing patients who survived from those who died after the surgery with an accurate risk estimation. Discrimination was confirmed for isolated valve surgery and CABG. The euroSCORE II overestimated the risk with a significant O/E ratio in isolated CABG. We observed an underestimation of the predictions in high-risk procedures. There was evidence of substantial heterogeneity between studies for parameters of discrimination and model predictive accuracy. At meta-regression analysis, study-level data considered as explanatory variables of euroSCORE II performance were not significant. However, subgroup analysis showed that the studies that evaluated 30-day mortality or included patients with a completely retrospective evaluation of euroSCORE II had a great impact on variability of the estimated measures. When hospital mortality was evaluated by including more recent procedures (patients operated after the euroSCORE II cohort), heterogeneity was not significant or less evident. We addressed the issue of heterogeneity between studies by computing pooled estimates from a random-effects model. Our data, despite the score performance differing across studies, confirmed the global validity of the euroSCORE II as a tool for patient selection. Part of the variability detected among studies may be due to differences in performance between centers or surgeons. D’Errigo and colleagues\textsuperscript{35} evaluated the amount of differences in adjusted mortality rates (30-day mortality after CABG) due to differences between centers with a significant proportion of variability explained by different performance among cardiac surgery centers. Kilic and colleagues\textsuperscript{30} considered the combined effect of volume of procedures performed for operative outcomes in mitral valve surgery: Both surgeon and hospital volume correlated significantly with operative mortality, and lower surgeon-volume persisted as a significant risk factor in the risk-adjusted analysis. Papachristofi and colleagues\textsuperscript{37} evaluated the impact of surgeons on outcome after cardiac surgery. They showed that in-hospital death, in a large cohort of cardiac surgical patients, was determined primarily by the patient risk profile and by a small but significant difference in outcome between surgeons. The number and range of participating units included in the development of the euroSCORE II were greater than those involved in the validation studies\textsuperscript{7-29} and fairly representative of current cardiac surgery to obtain a robust cardiac surgery risk model.\textsuperscript{3} Moreover, a small evidence of inter-hospital heterogeneity was detected in the euroSCORE II original cohort.\textsuperscript{3}

The performance of the recently updated euroSCORE II has been assessed in several studies.\textsuperscript{7-29} Most of the analyzed populations were retrospective or prospective for only the most recent cases. Large prospective national and regional multicenter registries including contemporary patients undergoing cardiac surgery have confirmed the good performance of the euroSCORE II as a generic model for all cardiac surgery.\textsuperscript{10,24} Grant and colleagues\textsuperscript{10} conducted the largest prospective study assessing the performance of the euroSCORE II in UK cardiac surgery analyzing the multicenter data from the Society for Cardiothoracic Surgery in Great Britain and Ireland Database that included 23,740 procedures between July 2010 and March 2011 from all National Health Service hospitals and some UK private centers. The euroSCORE II performed well with a poor calibration for isolated CABG surgery and in both the highest and lowest risk patients.\textsuperscript{10} A significant overprediction of mortality for isolated CABG was observed in the largest study included in our meta-analysis. The analysis of data from a consortium of 17

### TABLE 2. Meta-regression analysis

<table>
<thead>
<tr>
<th>AUC ROC</th>
<th>Coefficient</th>
<th>P value</th>
<th>Observed/expected ratio</th>
<th>Coefficient</th>
<th>P value</th>
<th>Observed-expected difference</th>
<th>Coefficient</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean age (y)</td>
<td>0.000</td>
<td>.984</td>
<td>0.023</td>
<td>.280</td>
<td>0.092</td>
<td>.172</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multicenter study</td>
<td>-0.018</td>
<td>.617</td>
<td>-0.113</td>
<td>.692</td>
<td>-0.212</td>
<td>.817</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Presence of patients operated before 2010</td>
<td>-0.027</td>
<td>.360</td>
<td>0.044</td>
<td>.853</td>
<td>0.233</td>
<td>.755</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proportion of isolated CABG</td>
<td>0.001</td>
<td>.986</td>
<td>-0.210</td>
<td>.527</td>
<td>-0.359</td>
<td>.722</td>
<td></td>
<td></td>
</tr>
<tr>
<td>In-hospital mortality as main outcome</td>
<td>0.003</td>
<td>.945</td>
<td>-0.209</td>
<td>.494</td>
<td>-0.722</td>
<td>.448</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\textit{AUC}, Area under the curve; \textit{CABG}, coronary artery bypass grafting; \textit{ROC}, receiver operating characteristic.
cardiac surgical centers providing cardiac surgery in the Commonwealth of Virginia (50,588 procedures with 40,871 isolated CABGs performed between 2003 and 2012) showed a large deviation between observed and predicted mortality. This study had a great impact on heterogeneity and overall estimation. By excluding this study from the meta-analysis, the euroSCORE II showed a good accuracy for isolated CABG (O/E ratio not significant) with a significant underestimation in high-risk procedures (data not shown). Compared with the performance in all cardiac surgery, the euroSCORE II showed poorer discrimination and lower calibration for nonelective and high-risk surgery. The poor performance of the euroSCORE II in emergency procedures was demonstrated in both prospective registries that validated the model. No risk model is perfect, nor do all risk factors appear in all models. It is possible that concomitant risk factors not incorporated in the euroSCORE II and the variation in outcomes between centers or surgeons may reduce the accuracy of predicted events mainly in patients at higher risk.

One of the possible purposes of the cardiac risk score is to help assess the quality of cardiac surgical care, the so-called benchmarking, by comparing observed mortality rates with the expected rates. A risk model constructed for benchmarking should be accurate in all risk groups, so that fair comparisons between providers can be made. The validity of euroSCORE II for benchmarking of healthcare providers should be better evaluated.

Study Limitations

Only studies published in English were selected. There is the possibility of publication bias because studies obtaining favorable results could be published more easily than studies with unfavorable results in terms of mortality compared with predicted risk. Studies that included patients before the actual introduction of the euroSCORE II may be less accurate because the risk score was retrospectively computed, and dynamic changes could be present in the patient characteristics, case-mix, and baseline risk relative of patients who underwent surgery over a wide study period before the development of the euroSCORE II. We evaluated the overall agreement between predicted and observed frequencies in each study comparing mortality with expected risk. We were not able to evaluate calibration within subcategory of risk for lack of data.

FIGURE 3. Forest plot of studies estimating the AUC ROC (A), O/E ratio, and O-E difference (B and C) in isolated CABG, isolated valve surgery, and combined CABG with valve procedures. The first author and reference as shown in Table 1. ROC, Receiver operating characteristic; CABG, coronary artery bypass grafting.
The performance of the euroSCORE II across the entire spectrum of risk and different procedures or combination of surgeries should be better evaluated in populations with possible suboptimal calibration. High-risk patients should be more clearly defined and the analysis based over a greater number of centers and patients than those involved in this meta-analysis. Moreover, the effects of performing CABG with a beating-heart technique (off-pump) in comparison with cardiopulmonary bypass (on-pump) are not clearly established in terms of risk-prediction models. A patient-level analysis of prospective patients undergoing cardiac surgery should be performed to address both discrimination and calibration of the euroSCORE II. A pooled analysis of individual participant data, evaluating each risk factor of the euroSCORE II, may allow a deeper assessment of model calibration and recalibration also adjusting for confounding factors not included in the score.

CONCLUSIONS

The results from this meta-analysis demonstrate the validity of the euroSCORE II as a risk model for contemporary practice. The score has good discrimination and is accurate in predicting operative mortality after cardiac surgery. New prospective studies directly focusing on the euroSCORE II should be started to estimate the performance across different surgical categories, to detect trends over time, and to evaluate the influence of changes in patient characteristics on risk score accuracy.

References


FIGURE E1. Baujat plot of studies estimating the AUC ROC (A), O/E ratio (B), and O-E difference (C). Group A included studies with a retrospective evaluation of euroSCORE II or that evaluated 30-day mortality as main outcome. Group B included studies that evaluated hospital mortality by including only patients undergoing operation after the euroSCORE II cohort. The number displayed is the reference. ROC, Receiver operating characteristic.