Background. The European System for Cardiac Operative Risk Evaluation (EuroSCORE) was developed to predict immediate outcomes after adult cardiac operations, but less than 30% of the cases used to develop this score were valve procedures. We studied EuroSCORE performance in valve procedures.

Methods. We performed a meta-analysis of published studies reporting the assessment of discriminatory power of the EuroSCORE by receiver operating characteristics (ROC) curve analysis in adult valve operations. A comparison of observed and predicted mortality rates was also performed.

Results. A literature search identified 37 potentially eligible studies, and 12 were selected for meta-analysis comprising 26,621 patients with 1250 events (mortality rate, 4.7%). Meta-analysis of these studies provided an average area under the curve (AUC) value of 0.730 (95% confidence interval [CI], 0.717 to 0.743). The same results were obtained when meta-analyses were performed separately in studies categorized on reliability of uncertainty estimation: in the seven studies reporting reliable uncertainty estimation (8175 patients with 358 events; mortality rate, 4.4%), the ROC curve provided an average AUC value of 0.724 (95% CI, 0.699 to 0.749). The five studies not reporting reliable uncertainty estimation (18,446 patients with 892 events; mortality rate, 4.8%) had an average AUC of 0.732 (95% CI, 0.717 to 0.747). We documented a constant trend to overpredict mortality by EuroSCORE, both in the additive and especially in the logistic form.

Conclusions. The EuroSCORE has low discrimination ability for valve surgery, and it sensibly overpredicts risk. Alternative risk scoring algorithms should be seriously considered.

Risk stratification in cardiac surgery has become increasingly important in current clinical practice because it can provide benchmarks for hospital performances and can be useful to provide an estimation of the risk related to the operation. The additive and logistic European System for Cardiac Operative Risk Evaluation (EuroSCORE) models are widely used as risk prediction tools in adult cardiac surgery, especially in Europe [1]. These models were based on more than 19,000 consecutive patients who underwent operations at 128 European surgical centers ending in 1995 [2]. At that time, however, coronary artery bypass grafting (CABG) was dominating among the adult cardiac interventions, and less than 30% of operations involved valves [1].

Nevertheless, the EuroSCORE has been widely used to predict risk in valve surgery [3, 4], and it has been recently used to select high-risk patients for transcatheter aortic valve procedures and to compare survival after standard or transcatheter aortic valve replacement [5, 6]. In this study, we proposed a systematic review to answer the question whether it can predict early mortality after valve operations with sufficient accuracy: we assessed EuroSCORE discriminatory power using receiver operating characteristics (ROC) curves analysis and compared observed and EuroSCORE-predicted mortality rates.

Material and Methods

To study EuroSCORE performance in valve operations, a meta-analysis was done following Meta-analysis Of Observational Studies in Epidemiology (MOOSE) [7] and Standards for Reporting of Diagnostic Accuracy (STARD) [8] guidelines. On March 11, 2009, two reviewers searched Medline (1950 through March 7, 2009), and PubMed (up to March 11, 2009), including electronic links to related articles. The text string used was the single word “EuroSCORE.”

To further reduce the probability of losing any major related study, an electronic search was performed of four major cardiothoracic surgery journals in the electronic format, Interactive CardioVascular and Thoracic Surgery, The Annals of Thoracic Surgery, The European Journal of Cardiothoracic Surgery, and The Journal of Thoracic and Cardiovascular Surgery (available at http://ats.ctsnetjournals.org/
The journals were searched from January 1999 to March 2009 for the single word “EuroSCORE” in the full text of all articles. The title of every article was considered first, then selected abstracts were searched to identify reports about risk prediction by the additive or logistic EuroSCORE, or heart valve operations. The full texts of these articles were retrieved and searched for in-hospital or 30-day mortality data and for mortality prediction by the EuroSCORE. Figure 1 reports the flow chart of paper selection.

Once papers were identified, the selection criterion for meta-analysis inclusion for each study was the presence of an assessment of the discriminatory power of the EuroSCORE (additive or logistic model, or both) by ROC analysis in valve operations, with or without concomitant procedures such as CABG. In addition, exclusion criteria were (1) studies reporting ROC analysis in valve operations with fewer than 10 events occurring in the study period, (2) evidence of duplicate publication, (3) confounded patient population (ie, when the ROC analysis included other patient categories such as isolated CABG and they could not be differentiated from valves), and (4) failure to report raw mortality data.

Studies selected for meta-analysis were then classified as (1) studies with reliable uncertainty estimation (group A), when the figure of merit defined as the area under the curve (AUC), was reported with a reliable dispersion parameter (standard error or 95% confidence interval [CI], or both); and (2) articles without reliable uncertainty estimation (group B), if only the AUC was reported and its dispersion parameter was missing or unreliable [8]. The dispersion variables of these last studies were estimated using an approximate formula, derived from the asymptotic form of the standard error for the estimate of a population proportion [9]. Meta-analysis models were developed for all articles, and dividing articles based on the presence of uncertainty estimation to verify whether our estimate of the AUC did not seem to influence the results of the meta-analysis. Because the purpose of this meta-analysis was to gain insight in the general discriminatory characteristics of the EuroSCORE models (ie, whether they can discriminate) we included only the AUC instead of building a complete summary ROC curve. Our analysis follows the work of McClish [10] and Zhou [11]. We assessed heterogeneity using a one-sided \( \chi^2 \) measure and estimated the overall performance using inverse variance weights, as described in Parolari and colleagues [12].

The 37 potentially eligible studies (Fig 1) were also searched for reporting the observed and predicted (by either additive or logistic model) mortality rates to assess the ratio between observed and expected mortality rates; in this case, subanalyses considering different categories of patients (eg, only valves, valves plus CABG, high-risk patients) were considered only when five or more studies reported data for different categories of patients. Observed and predicted mortality rates were compared with the \( \chi^2 \) test with continuity correction. Multiple comparisons calculations were done following the Holm method [13].

All calculations were done with University of Chicago ROC software (http://xray.bsd.uchicago.edu/krl/), Excel 2003 (Microsoft Inc, Redmond, WA) or Mathematica 6 software (Wolfram Research Inc, Champaign, IL).

Results

Literature Search and Article Selection

A detailed description of patient selection is reported in the Appendix*. From these selection criteria, we could identify 12 studies; of these, 7 were classified as group A studies, including 8175 patients with 358 events, for a mortality rate of 4.4% [3, 14–19]; and 5 as group B, including 18446 patients with 892 events, for a mortality rate of 4.8% [20–24] (Table 1).

Meta-Analysis

Meta-analysis of all the 12 studies (Fig 2) provided an average AUC value of 0.730 (95% CI, 0.717 to 0.743) and the \( Q \) statistic (21.9, \( p = 0.056 \)) showed a nonsignificant amount of heterogeneity. The same results were obtained when meta-analyses were performed separately on the studies categorized on reliability of uncertainty estimation. Group A studies provided an average AUC of 0.724 (95% CI, 0.699 to 0.749) and the \( Q \) statistic (7.71, \( p = 0.36 \)) showed very little evidence of heterogeneity. Likewise,

*See note at end of article regarding e-only Appendix.
<table>
<thead>
<tr>
<th>Studies, Year</th>
<th>Period</th>
<th>Hospitals</th>
<th>Patient Categories</th>
<th>CABG? (%)</th>
<th>High-Risk?</th>
<th>Model</th>
<th>Pts.</th>
<th>Events (%)</th>
<th>95% CI</th>
<th>Pred.</th>
<th>O/E</th>
<th>AUC</th>
<th>95% CI (SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Group A studies</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Au, 2007 [14]</td>
<td>11/99–7/05</td>
<td>Single center</td>
<td>Valve pts.</td>
<td>N/A</td>
<td>No</td>
<td>ADDI</td>
<td>1406</td>
<td>67</td>
<td>4.77 (3.74–6.05)</td>
<td>5.2</td>
<td>0.917</td>
<td>0.77</td>
<td>0.70–0.83 (N/A)</td>
</tr>
<tr>
<td>Florath, 2006 [15]</td>
<td>3/96–12/03</td>
<td>Single center</td>
<td>AVR ± CABG</td>
<td>Yes (41)</td>
<td>No</td>
<td>LOGI</td>
<td>2198</td>
<td>84</td>
<td>3.82 (3.08–4.47)</td>
<td>N/A</td>
<td>0.7</td>
<td>0.66–0.75 (NA)</td>
<td></td>
</tr>
<tr>
<td>Heikkinen, 2007 [3]</td>
<td>1/93–12/00</td>
<td>Single center</td>
<td>MVR (regurg) ± CABG</td>
<td>Yes (37)</td>
<td>No</td>
<td>ADDI</td>
<td>180</td>
<td>18</td>
<td>10 (6.2–15.6)</td>
<td>N/A</td>
<td>0.804</td>
<td>0.689–0.919 (0.059)</td>
<td></td>
</tr>
<tr>
<td>Langanay, 2006 [16]</td>
<td>1/00–12/04</td>
<td>Single center</td>
<td>AVR ± CABG</td>
<td>Yes (19)</td>
<td>Yes (&gt;80 y)</td>
<td>ADDI</td>
<td>442</td>
<td>33</td>
<td>7.47 (5.27–10.4)</td>
<td>N/A</td>
<td>0.75</td>
<td>0.65–0.86 (N/A)</td>
<td></td>
</tr>
<tr>
<td>Osswald, 2009 [17]</td>
<td>1/94–3/06</td>
<td>Single center</td>
<td>AVR only</td>
<td>No</td>
<td>No</td>
<td>ADDI</td>
<td>1545</td>
<td>34</td>
<td>2.2 (1.55–3.10)</td>
<td>6.1</td>
<td>0.361</td>
<td>0.677</td>
<td>0.606–0.748 (N/A)</td>
</tr>
<tr>
<td>Troupoulis, 2005 [18]</td>
<td>1/92–3/02</td>
<td>Single center</td>
<td>Valve ± CABG</td>
<td>Yes (45)</td>
<td>No</td>
<td>ADDI</td>
<td>1105</td>
<td>70</td>
<td>6.33 (5.00–7.98)</td>
<td>8</td>
<td>0.791</td>
<td>0.72</td>
<td>0.66–0.78 (N/A)</td>
</tr>
<tr>
<td>Van Gameren, 2008 [19]</td>
<td>1/03–1/07</td>
<td>Single center</td>
<td>Valve only</td>
<td>No</td>
<td>No</td>
<td>ADDI</td>
<td>904</td>
<td>25</td>
<td>2.77 (1.84–4.12)</td>
<td>5.3</td>
<td>0.523</td>
<td>0.77</td>
<td>0.67–0.87 (N/A)</td>
</tr>
<tr>
<td>Van Gameren, 2008 [19]</td>
<td>1/03–1/07</td>
<td>Single center</td>
<td>Valve ± CABG</td>
<td>Yes (all)</td>
<td>No</td>
<td>ADDI</td>
<td>395</td>
<td>27</td>
<td>6.84 (4.64–9.91)</td>
<td>6.4</td>
<td>1.069</td>
<td>0.71</td>
<td>0.62–0.80 (N/A)</td>
</tr>
<tr>
<td><strong>Group B studies</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Total</td>
<td>8175</td>
<td>358</td>
<td>4.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bhatti, 2006 [20]</td>
<td>4/02–3/04</td>
<td>Multicenter</td>
<td>Valves only</td>
<td>No</td>
<td>No</td>
<td>LOGI</td>
<td>1523</td>
<td>53</td>
<td>3.48 (2.64–4.56)</td>
<td>7.9</td>
<td>0.441</td>
<td>0.79</td>
<td>N/A</td>
</tr>
<tr>
<td>Bhatti, 2006 [20]</td>
<td>4/02–3/04</td>
<td>Multicenter</td>
<td>Valves ± CABG</td>
<td>Yes (all)</td>
<td>No</td>
<td>LOGI</td>
<td>984</td>
<td>71</td>
<td>7.22 (5.71–9.06)</td>
<td>9.6</td>
<td>0.752</td>
<td>0.73</td>
<td>N/A</td>
</tr>
<tr>
<td>Gummert, 2009 [21]</td>
<td>1/06–12/07</td>
<td>Multicenter</td>
<td>AVR only</td>
<td>No</td>
<td>No</td>
<td>LOGI</td>
<td>6305</td>
<td>245</td>
<td>3.89 (3.43–4.40)</td>
<td>7.03</td>
<td>0.48125</td>
<td>0.693</td>
<td>N/A</td>
</tr>
<tr>
<td>Karthik, 2004 [22]</td>
<td>4/97–3/02</td>
<td>Multicenter</td>
<td>AVR + CABG</td>
<td>Yes (all)</td>
<td>No</td>
<td>ADDI</td>
<td>1769</td>
<td>154</td>
<td>8.71 (7.45–10.1)</td>
<td>6.7</td>
<td>1.300</td>
<td>0.73</td>
<td>N/A</td>
</tr>
<tr>
<td>Roques, 2001 [23]</td>
<td>9/95–12/95</td>
<td>Multicenter</td>
<td>Valve ± CABG</td>
<td>Yes (21)</td>
<td>No</td>
<td>ADDI</td>
<td>5672</td>
<td>344</td>
<td>6.06 (5.46–6.73)</td>
<td>N/A</td>
<td>0.75</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>Xu, 2007 [24]</td>
<td>1/04–1/06</td>
<td>Single center</td>
<td>Valve ± CABG</td>
<td>Yes (8)</td>
<td>No</td>
<td>ADDI</td>
<td>2193</td>
<td>25</td>
<td>1.14 (0.75–1.70)</td>
<td>N/A</td>
<td>0.69</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td><strong>All studies included in meta-analysis</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Total</td>
<td>18,446</td>
<td>892</td>
<td>4.8</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

ADDI = additive EuroSCORE; AUC = area under the curve; AVR only = isolated aortic valve replacement; AVR = aortic valve replacement; AVR + CABG = aortic valve replacement with comcomitant CABG; AVR ± CABG = aortic valve replacement with or without comcomitant CABG; CABG = coronary artery bypass grafting; CI = confidence interval; EuroSCORE = European System for Cardiac Operative Risk Evaluation; LOGI = logistic EuroSCORE; MVR = mitral valve replacement or repair (non specified); N/A = not available; O/E = observed/expected ratio; Pred. = predicted (%) in-hospital mortality by EuroSCORE; Pts. = number of patients; SE = standard error of the mean; Valve pts. = valve procedures not otherwise specified; Valve + CABG = single, double, triple valve procedure with concomitant CABG; Valve ± CABG = single, double, triple valve procedure with or without concomitant CABG; Valves only = single, double, triple valve procedures without concomitant CABG.
group B studies had an average AUC of 0.732 (95% CI, 0.717 to 0.747).

We investigated whether we could propose an operation-type analysis, but unfortunately, the studies differed in too many ways to allow us to reach any reliable conclusion. Given that studies involving different types of operation (valves only without CABG, high-risk patients, aortic valves, mitral valves) produced homogeneous outcomes relative to the estimation uncertainties, reliable statements about potential differences in the performance of the EuroSCORE among them could not be made. This does not mean the EuroSCORE performs identically, but rather, that much larger experiments are necessary to achieve discrimination. For this reason we did not present a subcategory analysis of the data. However, there is no reason to expect a much better performance for any of the procedures involved, and therefore, such a study might be of little relevance from the clinical perspective.

Observed vs Expected Mortality Rates
Of the 37 potentially eligible studies, 24 reported the observed and predicted mortality rates for the additive or logistic model, or for both, in valve operations (Table 2

<table>
<thead>
<tr>
<th>EuroSCORE</th>
<th>Studies</th>
<th>Patients</th>
<th>Events</th>
<th>% 95% CI</th>
<th>Predicted % 95% CI</th>
<th>O/E Ratio</th>
<th>p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Additive</td>
<td>Valve pts.</td>
<td>20</td>
<td>10,387</td>
<td>619</td>
<td>5.96</td>
<td>5.52–6.44</td>
<td>698.8</td>
</tr>
<tr>
<td>Logistic</td>
<td>Valve pts.</td>
<td>22</td>
<td>18,010</td>
<td>925</td>
<td>5.14</td>
<td>4.82–5.47</td>
<td>1551.0</td>
</tr>
<tr>
<td>Additive</td>
<td>Valve pts.-HR</td>
<td>8</td>
<td>1369</td>
<td>124</td>
<td>9.06</td>
<td>7.62–10.73</td>
<td>130.1</td>
</tr>
<tr>
<td>Logistic</td>
<td>Valve pts.-HR</td>
<td>8</td>
<td>1527</td>
<td>128</td>
<td>8.38</td>
<td>7.06–9.91</td>
<td>234.8</td>
</tr>
<tr>
<td>Additive</td>
<td>Valves only</td>
<td>7</td>
<td>3956</td>
<td>147</td>
<td>3.72</td>
<td>3.16–4.37</td>
<td>273.9</td>
</tr>
<tr>
<td>Logistic</td>
<td>Valves only</td>
<td>8</td>
<td>11,708</td>
<td>442</td>
<td>3.78</td>
<td>3.44–4.14</td>
<td>988.8</td>
</tr>
<tr>
<td>Additive</td>
<td>AVR ± CABG</td>
<td>9</td>
<td>5457</td>
<td>345</td>
<td>6.32</td>
<td>5.70–7.00</td>
<td>391.8</td>
</tr>
<tr>
<td>Logistic</td>
<td>AVR ± CABG</td>
<td>11</td>
<td>12,628</td>
<td>639</td>
<td>5.06</td>
<td>4.69–5.47</td>
<td>1133.5</td>
</tr>
<tr>
<td>Additive</td>
<td>AVR only</td>
<td>5</td>
<td>3042</td>
<td>121</td>
<td>3.98</td>
<td>3.32–4.75</td>
<td>225</td>
</tr>
<tr>
<td>Logistic</td>
<td>AVR only</td>
<td>5</td>
<td>9258</td>
<td>362</td>
<td>3.91</td>
<td>3.53–4.33</td>
<td>809.4</td>
</tr>
</tbody>
</table>

* Nonsignificant after multiple comparison correction.

AUC = area under the curve; AVR only = isolated aortic valve replacement; AVR ± CABG = aortic valve replacement with or without concomitant CABG; CABG = coronary artery bypass; CI = confidence interval; EuroSCORE = European System for Cardiac Operative Risk Evaluation; Exp. = expected; HR = high-risk; O/E = observed/expected ratio; Obs. = observed; Pts. = number of patients; Valve pts. = every valve procedure with or without concomitant CABG; Valves only = single, double, triple valve procedures without concomitant CABG.
The following patient categories were analyzed because at least five studies were available that reported data about the observed and the expected mortality rates: (1) all patients undergoing valve operations with or without concomitant CABG; (2) high-risk patients undergoing valve operations with or without concomitant CABG; (3) patients undergoing valve surgery without concomitant CABG; (4) patients undergoing aortic valve replacement with or without concomitant CABG; and (5) patients undergoing isolated aortic valve replacement.

Overall, data show a constant and statistically significant propensity of the logistic EuroSCORE to overpredict mortality in all the five patient categories (even after multiple-comparisons correction), whereas the additive EuroSCORE overpredicts in three of five subcategories: all patients undergoing valve surgery with or without concomitant CABG (the only one nonsignificant after multiple comparisons correction), patients undergoing valve surgery without concomitant CABG, and patients undergoing isolated aortic valve replacement. In other words, the additive EuroSCORE does not seem to overestimate risk for cases of higher complexity (high-risk patients and valves plus CABG); on the other hand, the logistic EuroSCORE performs poorly in all subcategories of patients, constantly overpredicting mortality.

**Comment**

Scoring systems that allow perioperative risk estimation are important tools in medical quality control and in performance evaluation of hospitals and surgeons. Several risk scoring systems are currently used in adult cardiac surgery, and the EuroSCORE, both in its additive and logistic version, is the one most frequently used in Europe. This method, which requires a relatively simple data collection (17 items), has been shown to have relatively good discriminatory performance in adult cardiac operations in European [25] and North American patients [26]. It is still unknown, however, whether the best discriminatory performance by the EuroSCORE is achieved in case of CABG procedures [21, 27] or in valves [20, 28], and whether the differences in performance depend on the type of procedure. Given that at the time of its development in 1995 isolated CABG was by far the most common indication for adult cardiac operations—almost two-thirds of the procedures—whereas valve procedures accounted for little less than 30% of the procedures [2], some authors have raised questions about the reliability of this scoring system in valve procedures [19], especially in aortic interventions and in higher-risk patients [29, 30].

The discriminatory power and precision in risk prediction of the EuroSCORE in valve surgery has recently become increasingly important for two reasons. The first is that in the most centers, valve procedures—either isolated or combined—actually represent more than 50% of the total caseload; therefore, accurate risk estimation in this patient population—mainly elderly and very elderly people—has become much more important. The second reason is strictly related to the recent evolution in technical options in aortic valve operations that has led to a steady increase in the adoption of transcatheter aortic valve procedures in patients at the highest risk or in very elderly people [31]. A correct risk prediction is essential to select the patients who might benefit the most from this new and still experimental technique while at the same time managing effectively health program expenditures, because these new devices are quite expensive. Although some authors have already suggested that the EuroSCORE might be an effective tool for the selection of these patients [5, 6], the question of appropriateness of the use of these models must be addressed before such a decision process can be implemented [32].

This study has been designed mainly to define the role of the EuroSCORE in risk prediction for the current clinical cardiac surgical practice that progressively involves valve patients who are older. It was designed to attempt providing useful insights in the use of additive and logistic EuroSCORE models in defining which patients among the candidates to aortic valve replacement are at highest risk and therefore might be the best candidates for alternative and innovative transcatheter procedures.

Unfortunately, our study strongly suggests that the EuroSCORE might not be the appropriate tool for risk prediction in isolated valve operations or those combined with other cardiac procedures. The AUC derived from the meta-analysis provided estimates of 0.72 to 0.74, which are in a range of a performance considered less than satisfactory for a risk stratification algorithm [33]. EuroSCORE discrimination is also substantially lower with respect to the performance of the Society of Thoracic Surgery (STS) algorithm, which is about 0.8 for isolated valve operations [34] and about 0.75 for valves plus CABG [35]. The explanation for this is that the STS score is updated almost annually, and, for this reason, it may better follow the changes occurring in valve patient population with relative ease, whereas the EuroSCORE is now undergoing its first revision since its introduction.

That the heterogeneity test was almost significant is consistent with differences actually existing among the different procedures studies here, which should be expected, but given the large number of alternatives, it would require a large and detailed study to investigate it. Perhaps these resources would be better used to design a new more specific risk prediction scheme. Interestingly, a preliminary analysis of our data in terms of ROC analysis showed similar discriminatory performance for additive and logistic EuroSCORE models, suggesting a near equivalence for these two models from this perspective. In addition, the analysis of the ratio between observed and expected deaths by the additive and logistic models suggest overall a tendency for both models to significantly overpredict the mortality risk, which appears to be more marked for the logistic model.

In detail, the logistic EuroSCORE constantly and significantly overpredicted mortality in all five subcatego-


This appears to contrast with previous findings suggesting that the more recent logistic model is more appropriate for risk prediction, especially in more complex cases. In fact, it appears that we are observing a EuroSCORE paradox: the older and less sophisticated algorithm—the additive EuroSCORE—outperforms the more recent and complex method implemented just to estimate with more precision the risk in complicated cases. Our study adds further evidence that in current cardiac surgical practice, there is a compelling need for perhaps even a complete reengineering of EuroSCORE. Almost 15 years have elapsed since its development, and the clinical profile of cardiac surgical patients, especially for valve procedures, has changed substantially and so should the methods used to assess their risk.


*The Appendix is available only online. To access it, please visit [http://ats.ctsnetjournals.org](http://ats.ctsnetjournals.org) and search for the article by Parolari, Vol. 89, pages 787–93.
Appendix

Literature Search and Article Selection

The literature search identified 37 potentially eligible studies [1–37]; of these, 17 did not meet the mandatory inclusion criterion [4, 5, 7, 8, 10–13, 16, 22, 23, 25, 26, 31, 32, 34, 36], and 8 were excluded because of the presence of exclusion criteria (Fig 1) [2, 6, 18–20, 24, 28, 30].

Twelve studies were then selected for meta-analysis purposes [1, 3, 9, 14, 15, 17, 21, 27, 29, 33, 35, 37]; two studies [3, 35] reported the area under the curve (AUC) for patients undergoing valve operations and valve operations plus coronary artery bypass grafting (CABG) separately, and data concerning both these European System for Cardiac Operative Risk Evaluation (EuroSCORE) performances have been included in the meta-analysis. In addition, it should also be noted that among the articles selected for meta-analysis, five [15, 17, 27, 35, 37] reported the performance of both additive and logistic EuroSCORE on the same patients subsets (Appendix Table).

Interestingly, all the AUC values computed using the additive or the logistic models for the same cases produced essentially identical results [15, 17, 27, 35, 37], even when the analysis was done only for subcategories [35]. This result, consistent with our previous findings [38] suggests that for the purpose of assessing discriminating ability using the AUC, we can safely use the two types of measurement interchangeably. For this reason every study was used once, favoring the additive model (because there are larger number of these studies), with the addition of only one study that reported logistic, but not additive, EuroSCORE performance [9]. Finally, one article [37] provided confidence intervals of the AUC for the logistic and additive EuroSCORE that were unrealistically small. This was probably because they used a method for computing variances that is known to be biased low when samples are very unbalanced (very different number of positive and negative cases). We therefore considered the method as providing no estimate of measurement error and classified the study as low quality.

Observed vs Expected Mortality Rates

The studies that reported the observed and predicted mortality rates for the additive or logistic model, or for both, in valve operations were references 1, 3–5, 7, 8, 10–17, 20, 22, 25, 27–29, and 32–35.

Appendix Table. Studies Included in the Meta-Analysis Reporting EuroSCORE Performance for the Additive and Logistic Models

<table>
<thead>
<tr>
<th>Study, Year</th>
<th>Model</th>
<th>Pts.</th>
<th>Events (%)</th>
<th>95% CI Pred.</th>
<th>O/E</th>
<th>AUC</th>
<th>95% CI (SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Karthik, 2004 [17]</td>
<td>ADDI</td>
<td>1769</td>
<td>154</td>
<td>8.71 (7.45–10.1)</td>
<td>6.7</td>
<td>1.300</td>
<td>0.73 (N/A)</td>
</tr>
<tr>
<td>Karthik, 2004 [17]</td>
<td>LOGI</td>
<td>1769</td>
<td>154</td>
<td>8.71 (7.45–10.1)</td>
<td>9.4</td>
<td>0.927</td>
<td>0.73 (N/A)</td>
</tr>
<tr>
<td>Heikkinen, 2007 [15]</td>
<td>ADDI</td>
<td>180</td>
<td>10</td>
<td>18 (6.2–15.6)</td>
<td>5</td>
<td>2.000</td>
<td>0.804 (0.059)</td>
</tr>
<tr>
<td>Heikkinen, 2007 [15]</td>
<td>LOGI</td>
<td>180</td>
<td>10</td>
<td>18 (6.2–15.6)</td>
<td>3.7</td>
<td>2.703</td>
<td>0.806 (0.057)</td>
</tr>
<tr>
<td>Osswald, 2009 [27]</td>
<td>ADDI</td>
<td>1545</td>
<td>34</td>
<td>2.2 (1.55–3.10)</td>
<td>6.1</td>
<td>0.361</td>
<td>0.677 (N/A)</td>
</tr>
<tr>
<td>Osswald, 2009 [27]</td>
<td>LOGI</td>
<td>1545</td>
<td>34</td>
<td>2.2 (1.55–3.10)</td>
<td>9.3</td>
<td>0.237</td>
<td>0.666 (N/A)</td>
</tr>
<tr>
<td>Van Gameren, 2008 [35]</td>
<td>ADDI</td>
<td>904</td>
<td>25</td>
<td>2.77 (1.84–4.12)</td>
<td>5.3</td>
<td>0.523</td>
<td>0.67–0.87 (N/A)</td>
</tr>
<tr>
<td>Van Gameren, 2008 [35]</td>
<td>LOGI</td>
<td>904</td>
<td>25</td>
<td>2.77 (1.84–4.12)</td>
<td>6.1</td>
<td>0.454</td>
<td>0.66–0.86 (N/A)</td>
</tr>
<tr>
<td>Van Gameren, 2008 [35]</td>
<td>ADDI</td>
<td>395</td>
<td>27</td>
<td>6.84 (4.64–9.91)</td>
<td>6.4</td>
<td>1.069</td>
<td>0.71 (N/A)</td>
</tr>
<tr>
<td>Van Gameren, 2008 [35]</td>
<td>LOGI</td>
<td>395</td>
<td>27</td>
<td>6.84 (4.64–9.91)</td>
<td>7.8</td>
<td>0.877</td>
<td>0.72 (N/A)</td>
</tr>
<tr>
<td>Xu, 2007 [37]</td>
<td>ADDI</td>
<td>2193</td>
<td>25</td>
<td>1.14 (0.75–1.70)</td>
<td>1.14</td>
<td>0.71</td>
<td>0.69–0.73 (N/A)</td>
</tr>
</tbody>
</table>

ADDI = additive EuroSCORE; AUC = area under the curve; CI = confidence interval; EuroSCORE = European System for Cardiac Operative Risk Evaluation; LOGI = logistic EuroSCORE; N/A = not available; O/E = observed/expected ratio; Pred. = predicted (%) in-hospital mortality by EuroSCORE; Pts. = number of patients; SE = standard error of the mean.

References


