Institutional report - Cardiac general

Additive and logistic EuroSCORE performance in high risk patients

Ganesh Shanmugam, Mark West, Geoff Berg*

Abstract

We compared the performances of the additive and logistic EuroSCORE in predicting mortality in high-risk cardiac surgical patients, at a single institution. Both models were applied to 6535 patients, operated on at the Western Infirmary, Glasgow from March 1994 to August 2004. Calibration and discrimination were assessed using the Hosmer-Lemeshow (HL) Chi-square test and areas under the ROC curve. Overall mortality was 2.95%. Predicted mortalities were 4.1% (additive) and 5.2% (logistic). Actual mortality was 0.6% in the low risk (additive EuroSCORE 1-2), 2.1% in the medium risk (EuroSCORE 3-5) and 7% in the high-risk groups (EuroSCORE 6 plus). Actual mortality increased beyond a predicted risk of 8-10%. At the low risks both systems slightly over-estimated mortality, with the logistic EuroSCORE being more accurate. At EuroSCORES between 10-20, the additive EuroSCORE under-estimated risk, while the logistic EuroSCORE over-estimated mortality. Both systems were inaccurate at high risk. The HL statistics were 11.15 [P<0.04] for the additive and 37.78 [P<0.047] for the logistic models. ROC curve areas were 0.749 + 0.04 (additive) and 0.746 + 0.03 (logistic). The additive EuroSCORE model remains a simple system for cardiac risk assessment. The logistic EuroSCORE was not more accurate even in high-risk patients.

Keywords: EuroSCORE; Validation; Quality control; Risk stratification; Mortality

1. Introduction

The EuroSCORE is a prognostic scoring system developed in Europe for patients undergoing cardiac surgery. The evidence suggests that the additive EuroSCORE generally over-estimates mortality at lower EuroSCOREs (EuroSCORE ≤6) and under-estimates risk in high-risk patient groups EuroSCORE > 13 [1].

The performance of the logistic euroscore has not been extensively analyzed. Gogbashian and associates [1], in a systematic review of the international performance of the EuroSCORE, found no individual center studies that calculated the logistic EuroSCORE from their patient data. The only recently published study that evaluated both additive and logistic EuroSCOREs in high-risk patients was that of Michel et al. [2].

The objective of this study was to assess the performance of both models, especially in high-risk groups, in a large independent single center population of cardiac surgical patients.

2. Materials and methods

Both EuroSCORE models were applied to predict mortality in 6535 patients from the Western Infirmary database. The data from our institution were used and contributed to the development of the original EuroSCORE dataset. The dataset had been in use even previously. The data for all years were validated and are complete and accurate.

*Corresponding author. Tel.: +44 141 211 2807; fax: +44 141 211 1751. E-mail address: geoff.berg@northglasgow.scot.nhs.uk (G. Berg).

© 2005 Published by European Association for Cardio-Thoracic Surgery

The scoring system was first applied to three risk groups (0-2; 3-5 and >6). Absolute mortality was determined for the overall patient population and trends in actual mortality were analyzed, across the entire risk spectrum. On this basis, the overall population was split into broad subgroups. Risk prediction was calculated for the overall population and for these subgroups. Finally, we compared the performance of the two systems in high-risk patients.

The study population included 6535 patients operated upon at the Western Infirmary, between March 1994 and August 2004. The mean age for the overall population was 64.3 years. Repair of post-infarct ventricular septal defects and aneurysms was done in 33 patients. There were 181 redo procedures. On the basis of priority of care, operative procedures were categorized as being elective (55.6%), priority (27.9%), urgent (14%), emergent (2.2%) or salvage (0.3%).

3. Results

For the overall group, there were 193 deaths yielding an actual mortality of 2.95% [2.81–3.23]. The additive and logistic EuroSCOREs for the overall group were 4.1% [3.95–4.27] and 5.2% [4.97–5.46], respectively.

Table 1 and Fig. 1A summarizes the observed and expected mortality in the 3 conventional EuroSCORE sub-groups. Actual mortality was 0.6% [0.52–0.71] in the low risk (additive EuroSCORE 1–2), 2.1% [1.91–2.34] in the medium risk (EuroSCORE 3–5) and 7% [6.53–7.51] in the high-risk groups (EuroSCORE 6 plus). Both models appeared to over-estimate risk in all three categories.
Table 1
Observed and expected mortality in the 3 conventional EuroSCORE sub-groups. The figures in brackets indicate 95% confidence intervals

<table>
<thead>
<tr>
<th>Patient group</th>
<th>Number</th>
<th>Actual mortality % [95% CI]</th>
<th>Additive EuroSCORE % [95% CI]</th>
<th>Logistic EuroSCORE % [95% CI]</th>
<th>Actual additive ratio</th>
<th>Actual logistic ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>0–2</td>
<td>2210</td>
<td>0.6% [0.52–0.71]</td>
<td>1.1% [1.05–1.17]</td>
<td>1.3% [1.19–1.43]</td>
<td>0.55</td>
<td>0.46</td>
</tr>
<tr>
<td>3–5</td>
<td>2477</td>
<td>2.10 [1.91–2.34]</td>
<td>3.90 [3.79–4.01]</td>
<td>3.10 [2.81–3.39]</td>
<td>0.54</td>
<td>0.68</td>
</tr>
<tr>
<td>Total</td>
<td>6535</td>
<td>3.00 [2.81–3.23]</td>
<td>4.10 [3.95–4.27]</td>
<td>5.20 [4.97–5.46]</td>
<td>0.73</td>
<td>0.58</td>
</tr>
</tbody>
</table>

Fig. 1. A. Observed and expected mortality in the 3 conventional EuroSCORE sub-groups. B. Observed and predicted mortality in the EuroSCORE sub-groups. Reclassified on the basis of trends in actual mortality.

Table 2 summarizes the observed and expected mortality in different EuroSCORE sub-groups, on the basis of trends in actual mortality.

Fig. 2 displays trends in actual mortality across the entire risk spectrum.

Actual mortality remains low until the 8–10 risk category, rises up to a risk category of about 18. In higher risk categories, actual mortality increases sharply and exceeds 50%. On this basis, we re-stratified risk into 4 broad categories.

Fig. 3 shows the predicted mortality by subgroup (on the basis of additive EuroSCORE) plotted against (A) actual mortality (B) ADDITIVE euroscore and (C) LOGISTIC euroscore. This figure shows a slight over-estimation of mortality by additive and logistic EuroSCOREs at lower predictions. The crossover point at which the two systems begin to diverge is at a predicted mortality of between 8 and 10%. Beyond this additive EuroSCORE begins to underestimate mortality, but the corresponding curve stays fairly close to the actual mortality curve up to a risk of 18–20. The logistic EuroSCORE, on the other hand, continues to over-estimate mortality significantly as indicated by the wide divergence between the logistic and actual mortality curves.

Fig. 4 depicts the predicted mortality by subgroup plotted against (A) actual mortality (B) the difference between observed–predicted mortality (additive) and (C) the difference between observed–predicted mortality (logistic).

The degree of coherence between the mortality predicted by a scoring system and the actual mortality is determined by the proximity of the corresponding curve to line of neutrality (0).
Table 2
Observed and expected mortality in different EuroSCORE sub-groups, on the basis of trends in actual mortality

<table>
<thead>
<tr>
<th>Patient group</th>
<th>Actual mortality %</th>
<th>Additive EuroSCORE %</th>
<th>Logistic EuroSCORE %</th>
<th>Actual additive ratio</th>
<th>Actual logistic ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>0–10</td>
<td>2.40</td>
<td>3.80</td>
<td>3.90</td>
<td>0.63</td>
<td>0.62</td>
</tr>
<tr>
<td>11–15</td>
<td>15.60</td>
<td>11.70</td>
<td>29.90</td>
<td>1.33</td>
<td>0.52</td>
</tr>
<tr>
<td>16–20</td>
<td>19.50</td>
<td>17.40</td>
<td>69.60</td>
<td>1.12</td>
<td>0.28</td>
</tr>
<tr>
<td>21–24</td>
<td>50.00</td>
<td>21.00</td>
<td>89.00</td>
<td>2.38</td>
<td>0.56</td>
</tr>
</tbody>
</table>

The curves representing the difference between actual mortality and the additive and logistic EUROSCORES, respectively, lay in close proximity to the line of neutrality up to a risk of 8–10%.

Beyond this, the additive EuroSCORE underestimates mortality, but remains reasonably close to the actual mortality curve and the line of neutrality (0). The difference between actual mortality and predicted mortality (additive) is positive. The logistic EuroSCORE, however, continues to overestimate mortality in an increasing manner, as indicated by the corresponding logistic curve, which moves away from the line of neutrality. The difference between actual mortality and predicted mortality (logistic) is negative.

We then analyzed the performance of the two scoring systems, in relation to actual mortality in the high-risk group. Patients with risk exceeding 10% were assigned to this group, based on trends in actual mortality.

Table 3 summarizes the observed and expected mortality at high-risk EuroSCORE sub-groups. The additive euroscore underestimates risk, while the logistic euroscore overestimates actual mortality.

4. Calibration and discrimination

Calibration data by the Hosmer-Lemeshow Chi square test were 11.15 [P<0.47] for the additive and 37.78 [P<0.64] for the logistic models. To analyze the discriminatory power of the EuroSCORE on individual death prediction, areas under the ROC curves were measured (Fig. 5). Both systems
showed reasonable discrimination with values of 0.749 ± 0.04 [additive] and 0.746 ± 0.03 [logistic].

5. Discussion

Risk stratification delimits sub populations within a cohort that have different risks of a particular outcome. This knowledge assists in defining the indication for surgery and contributes to better resource allocation.

Without risk stratification, surgeons and hospitals treating high-risk patients will appear to have worse results than others. This may prejudice referral patterns, perpetuate a reluctance to operate on high-risk patients and affect resource allocation. This is undesirable because it is this group of patients, which stands to gain most from surgical treatment.

The EuroSCORE has been demonstrated to predict overall mortality reasonably well in many studies. The performance of the additive euroscore is supposedly less reliable in high-risk patients, while the role of the logistic EuroSCORE in this patient cohort is still largely undefined.

In comparison to the low and medium risk groups, the high-risk group (6–24) encompasses a wider spectrum and hence a large number of high risk patients need to be analyzed to render the findings statistically robust. The reduced number of events in the low-risk groups and the small sample sizes in the large risk groups induce statistical limitations. Gogbashian et al. [1] evaluated six large studies that evaluated the EuroSCORE, all of which had small patient numbers in the ‘high-risk’ category.

In order to assess additive EuroSCORE performance at higher scores it was therefore necessary to combine the data to produce significant numbers in order to produce a meaningful result. The conclusions were therefore less certain. Both Bridgewater et al. [3] and Sergeant et al. [4] have made this observation within their own data sets.

This is one of the largest single center studies evaluating the EuroSCORE. We had 1848 patients in the high-risk group – the largest number of high-risk patients, thus far in any single center study. The primary endpoint of this study was to evaluate and compare the two models, particularly in high-risk patients.

The concept of ‘high risk patients’ is differently defined in the literature [5–8]. The use of 6 as a cut-off point to identify high-risk patients is artificial [9]; an expected risk of 6 or higher seemed reasonable to select patients with an objectively increased mortality risk. An analysis of actual mortality trends in our study indicated that mortality remained low, up to a predicted EuroSCORE of 8–10.

The possibility of additive EuroSCORE to predict mortality in an individual center was explored by Sergeant et al. [4] who found that EuroSCORE overestimates mortality at low risk and underestimates it at high risk.

We analyzed over 6500 patients, and evaluated both scoring systems. Globally the observed mortality was lower than the predicted additive and logistic values. Our findings are in concordance with Sergeant et al. [4] in the risk group 0–8. Both scoring systems slightly overestimated mortality, with no significant difference between the two. The logistic EuroSCORE, however, overestimated mortality to a lesser extent than the additive EuroSCORE.

Beyond the 8–10 risk groups, additive EuroSCORE underestimates mortality. The logistic EuroSCORE on the other hand increasingly overestimates mortality up to a risk of 20. (Fig. 1B).

Intuitively, one would expect, that the logistic EuroSCORE would be more accurate, especially in high-risk groups, but we were unable to demonstrate this in our study.

Michel et al. [2], showed that both systems performed well with an area under the ROC curve of 0.783 (standard) and 0.785 (logistic).

Based on our findings, we were able to stratify patients into 3 broad and approximately equal risk groups:

1) 0–8 where actual mortality was low, predicted mortality was slightly higher and the logistic EuroSCORE was marginally more accurate. In this group, we would accept the overestimation of actual mortality and continue to rely on the additive EuroSCORE.

2) 8–16 where actual mortality increases significantly. The additive EuroSCORE underestimates mortality, but the logistic score overestimates it and was less accurate. It would be interesting to assess the performance of the EuroSCORE in this group, on a prospective basis.

3) >18–20, where actual mortality increases steeply, and both systems are inaccurate. In this group, the perfect scoring system does not exist, and therefore, risk scores should not be taken out of context while counseling patients, immaterial of their predictive accuracy. Clinical judgment is required for decision-making and informed consent.

The performance of the additive EuroSCORE across the entire risk spectrum is in keeping with findings in the other studies. We, however, failed to demonstrate that the logistic EuroSCORE was more accurate, in the high-risk patient group. However, this might probably be less a validation of the model, and more a variance of performance within this single center. Despite the large number of patients, this is a single center study. The large number of patients and the homogeneity of the population make the findings robust.
The fact that the simple additive score can be applied at the bedside without the use of sophisticated information technology and complex mathematics is a strong advantage. However, the combined impact of two or more risk factors on operative risk may be more than the simple sum of their parts, especially when each factor has an important impact on outcome. The logistic model, on the other hand, supposedly gives a more realistic prediction. For most cardiac surgical populations, this discrepancy has little impact on risk prediction as the numbers of such patients are very small and the differences seen are not of a magnitude that is sufficient to justify the complex calculations needed to work out the logistic EuroSCORE.

The overall acceptable prediction of the hospital mortality using the additive EuroSCORE is the consequence of an overestimation of the low risk and an underestimation of the high risk. A center effect can be the cause of the overestimation of the lower risk. This however applies only to the additive euroscore. Such an effect does not exist for the logistic euroscore, in this particular analysis.

6. Conclusion

In conclusion, the standard EuroSCORE continues to be a simple and easily applied system of risk assessment. The logistic EuroSCORE did not offer a distinct advantage over the standard model, even in high patients. Actual mortality begins to increase significantly, only above a predicted risk of 8–10%.

However, we continue to carry out risk stratification using both systems, due to the availability of on-site, accessible information technology, because high-risk surgery forms a substantial part of our workload, and our inherent interest in the application of risk modeling systems.

References