ELSEVIER

European Journal of Cardio-thoracic Surgery 36 (2009) 791—798

EUROPEAN JOURNAL OF CARDIO-THORACIC SURGERY

www.elsevier.com/locate/ejcts

An adjusted EuroSCORE model for high-risk cardiac patients

Marco Ranucci [a,*], Serenella Castelvecchio [a], Lorenzo A. Menicanti [a], Sabino Scolletta [b], Bonizella Biagioli [b], Pierpaolo Giomarelli [b]

[a] Department of Cardiothoracic — Vascular Anesthesia and Intensive Care, IRCCS Policlinico S. Donato, Milan, Italy
[b] Department of Surgery and Bioengineering, Unit of Cardiothoracic Surgery, University of Siena, Italy

Received 6 November 2008; received in revised form 8 February 2009; accepted 9 February 2009; Available online 8 April 2009

Abstract

Objective: To verify the accuracy and precision of the logistic European system for cardiac operative risk evaluation (EuroSCORE) in high-risk cardiac surgery patients and to develop and externally validate a new system of recalibration. **Methods:** The development series included 4279 high-risk patients who had undergone cardiac operations at the IRCCS Policlinico S. Donato. Performance, accuracy, and precision of the logistic EuroSCORE were assessed in this series, using a deciles-based comparison between expected and observed mortality rates, a receiver operating characteristic analysis, and a Hosmer—Lemeshow test for calibration. Differences between predicted and observed mortality rates were mathematically evaluated to develop an adjusted logistic EuroSCORE. This adjusted risk score was subsequently validated with the same approach on an external series of 1459 high-risk patients who had undergone cardiac operations at the Siena hospital. **Results:** The adjusted logistic EuroSCORE was based on five different correction factors applied to the crude logistic EuroSCORE depending on its value. At the external validation, this model provided a good performance, with observed mortality rates not significantly different from the expected in 8 out of 10 deciles of risk distribution. The adjusted EuroSCORE had the same moderate balanced accuracy of the crude logistic EuroSCORE (area under the curve: 0.695), with a better precision (Hosmer—Lemeshow calibration test: χ^2: 3.6, $p = 0.891$). **Conclusions:** Recalibration of the logistic EuroSCORE in high-risk patients is needed due to its tendency to overestimate the mortality risk. The application of a variable correction factor results in a better performance, increased precision, with unaltered balanced accuracy.

Keywords: Risk-adjustment; EuroSCORE; Logistic model; Cardiac surgery

1. Introduction

The European system for cardiac operative risk evaluation (EuroSCORE) has been developed in Europe to predict mortality risk for patients undergoing cardiac surgery [1,2]. The additive model first, and the more sophisticated logistic version later have been widely used and extensively tested worldwide, including individual European countries, North America, Japan, and Australia [3—8]. When externally validated, the two models have been considered accurate and precise by some authors [5—7,9]. However, in more recent years, concerns have been raised about the performance of both the additive and the logistic EuroSCORE (logES) [8]. Comparative studies have either failed to detect significant differences between the additive and the logistic model [12] or more often recognized a superiority to the logistic model [11,13] namely if applied to high-risk patients [4,10]. However, the majority of the authors stress that the logistic EuroSCORE tends to overestimate the expected mortality risk [8,10—12]. Several limitations must be considered when we test the ability of these models in the risk stratification of a single institution or, more importantly, in comparing hospitals or individual surgeon outcomes. Epidemiological factors, including changes in different risk factors profile over time, the inability of EuroSCORE to provide a discriminatory weighting for different operations and advances in surgical, anesthetic, and intensive care, in respect of intercenter differences, make a model introduced in 1999 unlikely to remain accurate and precise. Given the importance of this issue, different approaches have been proposed; however, whether physicians should use a ready-made model (developed in a similar population without any change) or recalibrate it on their own population or develop a new model using new variables coming from their own data is still an object of debate [14].

Aims of this study are (a) to verify the performance, accuracy, and precision of the logistic EuroSCORE in high-risk patients; (b) to introduce a new system of recalibration; and (c) to externally validate the recalibrated model in a different subset of patients.

* Corresponding author. Address: Department of Anesthesia and Intensive Care, IRCCS Policlinico S. Donato, Via Morandi 30, 20097 San Donato Milanese, Milan, Italy. Tel.: +39 02 52774320; fax: +39 02 55602262.
E-mail address: cardioanestesia@virgilio.it (M. Ranucci).

1010-7940/$ — see front matter
doi:10.1016/j.ejcts.2009.02.023

2. Patients and methods

The study population comprised two cohorts of patients: one is represented by the activity in the San Donato Hospital (years 2001–2007), 11,150 adult patients; the second is a series of 2411 consecutive patients (years 2003–2007) operated at the Siena Hospital. During the study period, in both hospitals the patients were operated by the same surgical teams, comprising five first operating surgeons in San Donato and two first operating surgeons in Siena. The two institutions were independent for surgical, anesthesia, and intensive care unit teams.

The San Donato cohort (development series) was used to assess the logES in terms of performance, significance, accuracy, and precision in the definition of the actual mortality risk. From the same cohort a subgroup of high-risk patients was extracted and re-analyzed. The relationship between expected and actual mortality rates was explored using a logistic regression analysis having the logES as independent variable and the mortality (0 = survived; 1 = dead) as dependent variables. Discrepancies between expected and observed mortality rates were explored by sorting the population according to the logES, and subsequently dividing it into 10 deciles of distribution of the logES. Deciles from 1 to 10 therefore represent incremental risk classes; each one containing about 10% of the population. For each decile, differences between expected and observed mortality rates were explored by comparing the mean values (with 95% confidence intervals).

Adequate mathematical calculations including a non-linear regression analysis having the logES as independent variable and the differences between expected and observed mortality rates as dependent variable have been applied to create an adjusted logistic EuroSCORE (ADlogES) model. Approximations to make the model clinically applicable have been introduced.

The adjusted model was tested for external validation in the Department of Cardiac Surgery of a different hospital (Policlinico Le Scotte, Siena, Italy). Differences between the San Donato and the Siena cohorts were explored for comorbidities and pathological characteristics.

The adjusted model was tested on the Siena cohort (validation series), after extracting a high-risk patient population. The new model was explored for significance, clinical performance (difference between expected and observed mortality rates), calibration (precision) and discrimination (balanced accuracy), and compared to the conventional, unadjusted model.

Additional tests included logistic regression analysis level of significance; Hosmer–Lemeshow test for calibration of the models; receiver operating characteristics (ROC) analysis with calculation of the area under the curve (AUC) for balanced accuracy of the models and non-linear regression analyses.

Subgroup analyses according to pathology and elective/emergency operation were performed.

A p value <0.05 was considered significant for all the statistical tests. Statistical calculations were performed using a computerized statistical program (SPSS 11.0, Chicago, IL).

3. Results

The San Donato cohort (development series) included 11,150 patients constituting the initial study population. The logES was tested as an independent predictor of hospital mortality in terms of statistical significance of the model, clinical performance, calibration, and discriminative power. The expected mortality was 6.6% (95% confidence interval 6.4–6.8) significantly ($p < 0.001$) higher than the observed hospital mortality (3.8%, 95% confidence interval 3.4–4.2). The relationship between the logES and the hospital mortality is defined by a logistic equation:

$$\text{Hospital mortality} = \frac{e^{-3.8+0.061\,\text{logES}}}{1+e^{-3.8+0.061\,\text{logES}}} \qquad \text{with a } p \text{ value} < 0.001.$$

This model has an overall good discriminative power, with an AUC of 0.763 (95% confidence interval 0.739–0.788), but demonstrated a poor calibration (Hosmer–Lemeshow χ^2: 59, $p < 0.001$).

The relationship between expected and observed mortality was analyzed (Fig. 1) for each decile of distribution of the logES. The graphical relationship shows a good fit for the first six deciles. This good fit was confirmed by the statistical comparison of expected versus observed mortality within each decile. There was no significant difference for the first, second, third, and sixth decile, whereas the logES significantly overestimated the mortality risk in the fourth and fifth deciles. From the seventh decile, the mortality risk was always overestimated by the logES. The sixth decile corresponds to a logES of 4.27, the seventh to a logES of 5.47. Limiting the analysis to the first six deciles (low to moderate mortality risk) the calibration improved and the Hosmer–Lemeshow χ^2 is 11, with a $p = 0.206$. We therefore defined as 'good fit area' that belonging to the first six deciles.

According to the deciles distribution, the population of patients having a logES >5.0 was defined as 'high-risk

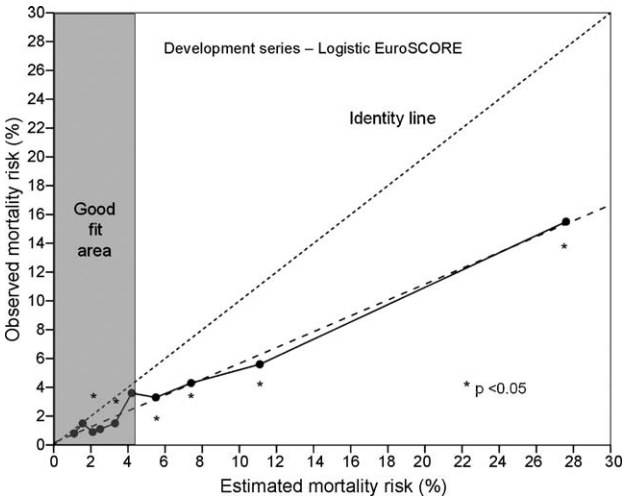

Fig. 1. Observed versus estimated mortality in the development series (all patients, $n = 11,150$).

patients', and separately analyzed for the purposes of the study.

The subsequent analysis aimed to assess an adjustment factor for the EuroSCORE is therefore intended to be dedicated to patients with a logistic EuroSCORE higher than 5.0.

This high-risk patient population was comprised of 4279 patients, and was explored following the same steps as above.

The expected mortality was 13.2% (95% confidence interval 12.8–13.5) significantly ($p < 0.001$) higher than the observed hospital mortality (7.4%, 95% confidence interval 6.6–8.2). The relationship between the logES and the hospital mortality is defined by a logistic equation:

$$\text{Hospital mortality} = \frac{e^{-3.28+0.046\,\text{logES}}}{1+e^{-3.28+0.046\,\text{logES}}} \quad \text{with a } p \text{ value} < 0.001. \qquad \text{(a)}$$

This model has an overall moderate discriminative power, with an AUC of 0.696 (95% confidence interval 0.664–0.728), and demonstrated a very poor calibration (Hosmer–Lemeshow χ^2: 22, $p = 0.006$).

The relationship between expected and observed mortality was analyzed (Fig. 2) for each decile of distribution of the logES. With the only exception of the second decile, the logES constantly and significantly overestimated the mortality risk. As a result, the linear regression approximating the expected mortality rate for each decile is consistently and progressively diverging from the identity line (perfect fit).

In order to create an adjustment function to correct this overestimation, the following steps were applied:

(a) For each possible value of expected mortality rate (logES) the true probability of death as established from the logistic regression Eq. (a) was calculated.
(b) For each value of logES the difference between the expected mortality rate and the true probability of death was calculated.

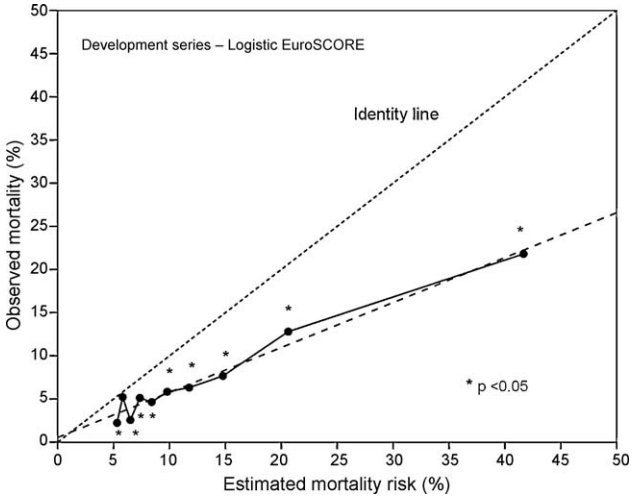

Fig. 2. Observed versus estimated mortality in the development series (high-risk patients, $n = 4279$).

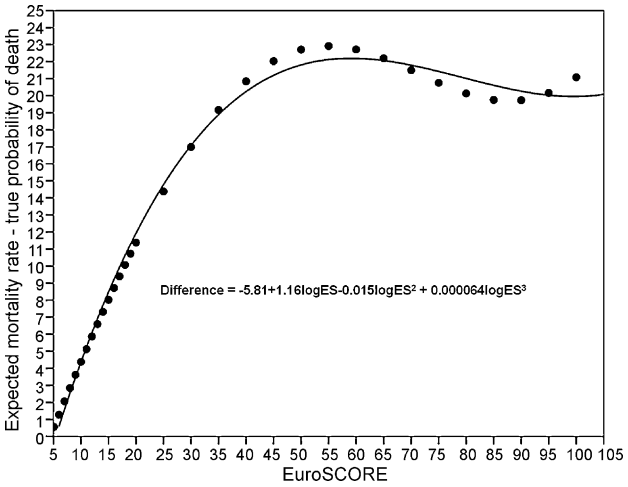

Fig. 3. Adjustment equation for the logistic EuroSCORE in the high-risk patient series.

(c) The difference (dependent variable) was plotted against the correspondent logES value (independent variable) (Fig. 3).
(d) The relationship was analyzed with a regression model for non-linear relationships. The best fit was found with a cubic equation ($r^2 = 0.994$, $p < 0.001$) reported in the figure.
(e) This cubic equation was considered the adjustment function for Eq. (a).

According to this adjustment function, the adjusted logistic EuroSCORE (ADlogES) would be:

$$\text{ADlogES} = \text{logES} + 5.81 - 1.16\text{logES} + 0.015\text{logES}^2 - 0.000064\text{logES}^3$$

The above adjustment equation, that is mathematically correct, but practically unuseful due to the complex calculations required, has been subsequently simplified with separate linear approximations, finally producing the practical adjustment model for grouped values of logES >5.0 reported in Table 1.

According to these adjusted values, our patient population was re-assessed with the ADlogES, and this value was subsequently tested in a new model, following the usual technique.

In this adjusted model, the expected mortality was 7.25% (95% confidence interval 7.1–7.4), not significantly different than the observed hospital mortality (7.4%, 95% confidence interval 6.6–8.2). The relationship between the ADlogES and

Table 1
Adjustment factors for grouped values of logES.

Baseline logES range	Adjustment factor	Range or values for ADlogES
5.1–6.0	ADlogES = 0.4 logES	2.0–2.4
6.1–25	ADlogES = 0.6 logES	3.6–15.0
25.1–40	ADlogES = 18	18.0
40.1–60	ADlogES = 20	20.0
>60	ADlogES = 30	30.0

ADlogES: adjusted logistic EuroSCORE.

Table 2
Patient profile in the development (San Donato) and validation (Siena) series.

EuroSCORE factor	Development (n = 4279), number (%) or mean ± SD	Validation (n = 1459), number (%) or mean ± SD	p
Age (years)	72.1 ± 10.3	70.4 ± 9.1	0.001
Gender female	1883 (44)	649 (44.5)	n.s.
Left ventricular ejection fraction	47.8 ± 13.1	51.1 ± 11.2	0.001
Serum creatinine value (mg/dl)	1.34 ± 1.00	1.20 ± 0.81	0.001
Recent myocardial infarction	1027 (24)	441 (30.2)	0.001
Chronic obstructive pulmonary disease	556 (13)	317 (21.7)	0.001
Neurologic dysfunction	513 (12)	117 (8)	0.001
Redo operations	513 (12)	189 (13)	n.s.
Emergency operations	372 (8.7)	185 (12.7)	0.001
Pulmonary hypertension	214 (5)	77 (5.3)	n.s.
Critical preoperative conditions	107 (2.5)	90 (6.2)	n.s.
Active endocarditis	64 (1.5)	41 (2.8)	0.002
Peripheral arteriopathy	787 (18.4)	265 (18.2)	n.s.
Other than isolated coronary operation	2798 (65.4)	935 (64.1)	n.s.
Thoracic aorta operation	321 (7.5)	98 (6.7)	n.s.
Post-myocardial infarction VSD repair	13 (0.3)	9 (0.6)	n.s.

SD: standard deviation; and VSD: ventricular septal defect.

the hospital mortality is defined by a logistic equation:

$$\text{Hospital mortality} = \frac{e^{-3.53+0.113\,\text{ADlogES}}}{1+e^{-3.53+0.113\,\text{ADlogES}}} \quad \text{with a } p \text{ value} < 0.001.$$

This model has the same discriminative power of the unadjusted one, with an AUC of 0.696 (95% confidence interval 0.664–0.728), but demonstrated a good calibration (Hosmer–Lemeshow χ^2: 12.3, p = 0.138).

The relationship between expected and observed mortality was re-analyzed (Fig. 4) in the new model for each decile of distribution of the ADlogES. In this analysis, the linear regression of the relationship between the expected and the observed mortality is very close to the identity line (observed mortality = −0.1 + 1.04ADlogES, r^2 = 0.955, $p < 0.001$). Moreover, in 9 deciles out of 10 the observed mortality rate was not significantly different from the expected one. In the second decile the observed mortality rate was significantly ($p < 0.05$) higher than the expected one (5.22%, 95% confidence interval 3.1–7.4, vs 2.52%, 95% confidence interval 2.5–2.6).

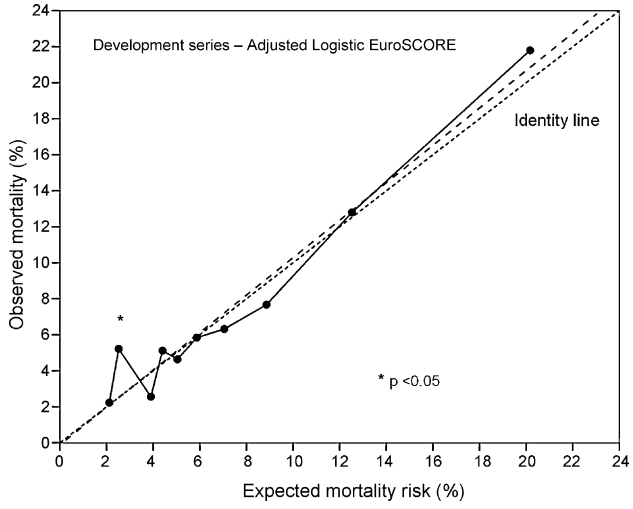

Fig. 4. Observed versus estimated mortality in the development series (high-risk patients) after adjustment of the logistic EuroSCORE.

This adjusted model was therefore considered both statistically and clinically acceptable for performance, discrimination, and calibration. It was therefore admitted to the subsequent external validation.

The Siena cohort (validation series) comprised a total of 2411 patients. From this series, a subset of high-risk patients (logES > 5.0) was extracted to match the same population of the test series.

This subgroup of high-risk patients included 1459 patients and constituted the final validation series. Table 2 presents the patient profile in the development and validation series. The validation series had a higher-risk profile, basically due to a higher rate of chronic obstructive pulmonary disease, emergency operations, critical preoperative conditions, active endocarditis. Non-significant higher rates of redo operations and post-myocardial infarction ventricular septal defect repair operations were present in the validation series. Age, ejection fraction, creatinine values, and rate of neurologic dysfunction were conversely worse in the development series.

Within the validation series, the logES was tested for relationship with the observed mortality following the usual technique based on a global and deciles-based analysis. In this model, the expected mortality was 17% (95% confidence interval 16.2–17.7), significantly ($p < 0.001$) higher than the observed hospital mortality (6.4%, 95% confidence interval 5.2–7.7). The relationship between the logES and the hospital mortality is defined by a logistic equation:

$$\text{Hospital mortality} = \frac{e^{-3.49+0.038\,\text{logES}}}{1+e^{-3.49+0.038\,\text{logES}}} \quad \text{with a } p \text{ value} < 0.001.$$

This model has a moderate discriminative power, with an AUC of 0.699 (95% confidence interval 0.64–0.75), and demonstrated a good calibration (Hosmer–Lemeshow χ^2: 3.93, p = 0.863). However, the overall performance is poor, with a regression line for the expected mortality in each decile that is diverging from the identity line (Fig. 5A) and in 8 deciles out of 10, the expected mortality was significantly ($p < 0.05$) overestimating the observed mortality.

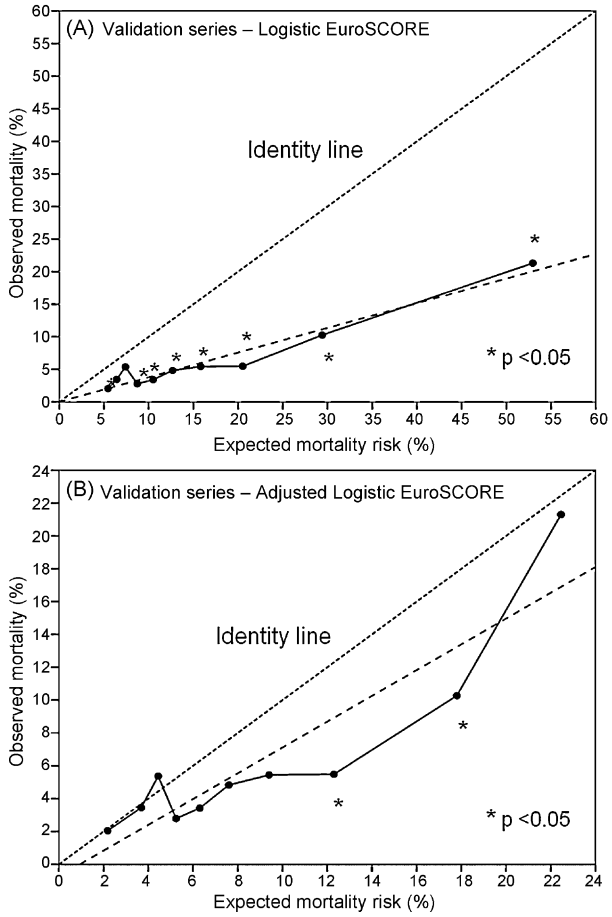

Fig. 5. Observed versus estimated mortality in the validation series (high-risk patients, $n = 1459$), before (A) and after (B) adjustment of the logistic EuroSCORE.

The same series was re-assessed using the ADlogES, and the new model was tested for validation.

The expected mortality on the basis of the ADlogES was 9.2% (95% confidence interval 8.9–9.5), still significantly ($p < 0.05$) higher than the observed mortality, but with a better approximation than the unadjusted model.

The relationship between the ADlogES and the hospital mortality is defined by a logistic equation:

$$\text{Hospital mortality} = \frac{e^{-3.77+0.098\,\text{ADlogES}}}{1+e^{-3.77+0.098\,\text{ADlogES}}} \quad \text{with a } p\text{ value} < 0.001.$$

This model has the same moderate discriminative power of the unadjusted one, with an AUC of 0.695 (95% confidence interval 0.64–0.75), and demonstrated an even better calibration (Hosmer–Lemeshow χ^2: 3.6, $p = 0.891$). When analyzed with a deciles-based approach, this adjusted model demonstrated no significant difference between expected and observed mortality rates in 8 deciles out of 10, with a significantly ($p < 0.05$) lower observed mortality rate in the eighth and ninth deciles (Fig. 5B). The linear regression line (observed mortality = 0.785 − 0.736ADlogES, $r^2 = 0.821$, $p < 0.001$) is closer to the identity line than the unadjusted model.

Table 3
Subgroup analysis for mortality in the overall population (development and validation series).

Group	Expected (ADlogES)% (95% CI)	Observed% (95% CI)	Ratio	p
Overall	7.7 (7.6–7.9)	7.2 (6.5–7.8)	1.07	n.s.
Isolated CABG	7.0 (6.6–7.1)	6.3 (5.3–7.8)	1.11	n.s.
Isolated valve	7.6 (7.0–7.7)	5.1 (3.4–6.5)	1.49	<0.05
CABG + valve	8.5 (8.0–8.7)	10.6 (8.8–12.4)	0.8	<0.05
Elective	6.9 (6.6–7.1)	6.1 (5.4–7.6)	1.13	n.s.
Emergency	14.8 (13.6–15.9)	17.9 (13.9–21.8)	0.83	n.s.

ADlogES: adjusted logistic EuroSCORE; CABG: coronary artery bypass graft; and CI: confidence interval.

To test for patient subgroup differences between the expected and observed mortality values using the ADlogES, the overall population (development and validation series) was investigated separately for pathology and operation characteristics (elective or emergency) (Table 3).

In the overall population there was no difference between expected and observed values. No difference was found in isolated coronary operations, elective operations, emergency operations; a significant overestimation of mortality risk was detected in isolated valve procedures and a significant underestimation in combined coronary–valvular procedures.

4. Discussion

There is a general agreement about the need for recalibrating the EuroSCORE, and namely the logistic EuroSCORE for high-risk patients, in order to obtain a mortality risk prediction more adherent to the daily clinical practice. There is no doubt that the logistic EuroSCORE, which works better than the additive one in the subset of high-risk patients, results in an overestimation of the mortality risk. This overestimation varies according to different series and different subsets of operation; however, the majority of the authors suggested a recalibration factor in the range of 0.4–0.5 [8,10,11,13,15] or up to 0.76 for more complex surgeries [11].

We must highlight the possibility that the constant overestimation of mortality risk may result from a publication bias: overperforming institutions are of course more likely to publish their results, and we cannot exclude that underperforming institutions are not willing to publish their data. However, with the present knowledge, an adjustment factor aimed to reduce the predicted mortality risk in high-risk patients seems necessary.

In our study we propose a variable correction factor to be applied to the logES in high-risk patients.

This correction factor is intended to adjust the mortality risk prediction in patients with a logES higher than 5.0. Patients with a lower logES do not need any adjustment, according to our evaluation.

The ADlogES has a better clinical performance than the logES when both internally and externally validated, with unchanged discrimination power and better calibration. The main novelty of this approach is the recalibration of the logES based on different factors according to the logES itself, with a resulting adjusted model that includes either variable

recalibration factors (0.4 for logES between 5.1 and 6.0, and 0.6 for logES between 6.1 and 25) or fixed values of ADlogES in very high-risk patients (18, 20, or 30 for logES respectively between 25.1 and 40, 40.1 and 60, or higher than 60). The resulting recalibrated model is easy to use and does not require complex calculations.

To address the meaning of a recalibration process of the EuroSCORE we should consider the statistical and clinical value of at least three different concepts: the performance, the discrimination (balanced accuracy), and the precision (calibration) of a risk model.

The most clinically relevant variable is the performance. If we define it as 'how well' a model (logES or ADlogES) accounts for the outcome (mortality), the best way for addressing it is looking at the expected versus observed mortality rates (with 95% confidence intervals) in the overall population, and in subsets with increasing mortality risk (arbitrarily defined as low, medium, high, or based on quintiles–deciles distribution). If a model is performing well, the expected mortality in the overall population and in the majority of the subgroups will not be significantly different from the observed one. This is not the case of the logES in high-risk patients; previous studies have demonstrated its overestimation, and the present study confirms its poor clinical performance and its overestimation of mortality risk.

However, once decided on a recalibration of the model, the other two parameters should be carefully considered. Balanced accuracy of the model, as defined by the AUC at the ROC analysis, will not change in any case, despite the use of fixed or variable recalibration factors. To modify it, only a totally new logistic model (which could or could not include the same variables of the EuroSCORE) should be developed. Therefore, a prerequisite for a recalibration process is that the overall balanced accuracy of the underlying logES model is satisfying (condition verified by many previous studies and by the presently analyzed data).

By applying a fixed recalibration factor, i.e. 0.4 as proposed by some authors [15] the clinical performance of the model will improve. However, the precision of the model (calibration), that simply means how close the values are to each other, will not change at all (the Hosmer–Lemeshow χ^2 will be unchanged). If we want to improve the clinical performance *and* the precision of the model, a variable recalibration factor is needed, according to different values of logES.

The problem of using risk models in clinical practice was nicely addressed by Ivanov and co-workers about 10 years ago [14]. These authors basically tested three options: (a) using a ready-made risk model; (b) recalibrating it by retesting the same risk factors and attributing different weights to each one; and (c) totally redoing the risk analysis and remodeling a new risk score. In their analysis, the accuracy of the models was practically the same, but the ready-made model had a very poor precision, opposite to an acceptable precision for both the recalibrated and remodeled risk scores.

In our adjusted model for high-risk patients, we have applied a recalibration strategy, ending up with an ADlogES that has the same balanced accuracy, a better precision, and a highly superior clinical performance than the logES. When applied to subgroups of patients, this adjusted model demonstrated an accurate performance regardless of the elective/emergency nature of the operation, and in isolated coronary operations. However, a significant overestimation of mortality in isolated valve procedures and underestimation in combined coronary–valve procedures was observed and should be considered for future studies. Our recalibration strategy was however different from the one proposed by Ivanov and co-workers: instead of re-analyzing every risk factor and reweighting them, we applied different correction factors to the logES, based on the crude value of the logES itself.

Predictive models are known to perform well when applied to the development series, but usually have a lower performance when applied to different series in different locations or even in different periods of time [14]. This of course applies to recalibrated models as well. The main strength of the present study is that the recalibrated model was externally validated in a different patient series, maintaining its basic properties in terms of performance, accuracy, and precision.

A limitation of this study is the retrospective nature of data collection and the long period of time explored. We admit that prospective studies excluding patients treated more than 5 years ago may offer additional advantages.

The recalibration is based upon complex calculations, but does not require complex calculations: being easy to apply; therefore, the ADlogES may be a useful clinical tool for stratifying mortality risk in cardiac operations.

References

[1] Roques F, Nashef SA, Michel P, Gauducheau E, de Vincentiis C, Baudet E, Cortina J, David M, Faichney A, Gabrielle F, Gams E, Harjula A, Jones MT, Pintor PP, Salamon R, Thulin L. Risk factors and outcome in European cardiac surgery: analysis of the EuroSCORE multinational database of 19030 patients. Eur J Cardiothorac Surg 1999;15:816–22.

[2] Nashef SA, Roques F, Michel P, Gauducheau E, Lemeshow S, Salamon R. European system for cardiac operative risk evaluation (EuroSCORE). Eur J Cardiothorac Surg 1999;16:9–13.

[3] Roques F, Michel P, Goldstone AR, Nashef SA. The logistic EuroSCORE. Eur Heart J 2003;24:881–2.

[4] Michel P, Roques F, Nashef SA, EuroSCORE Project Group. Logistic or additive EuroSCORE for high-risk patients? Eur J Cardiothorac Surg 2003;23:684–7.

[5] Roques F, Nashef SA, Michel P, Pinna Pintor P, David M, Baudet E, The EuroSCORE Study Group. Does EuroSCORE work in individual European countries? Eur J Cardiothorac Surg 2000;18:27–30.

[6] Nashef SA, Roques F, Hammill BG, Peterson ED, Michel P, Grover FL, Wyse RK, Ferguson TB, EuroSCORE Project Group. Validation of European System for Cardiac Operative Risk Evaluation (EuroSCORE) in North American cardiac surgery. Eur J Cardiothorac Surg 2002;22:101–5.

[7] Kawachi Y, Nakashima A, Toshima Y, Arinaga K, Kawano H. Risk stratification analysis of operative mortality in heart and thoracic aorta surgery: comparison between Parsonnet and EuroSCORE additive model. Eur J Cardiothorac Surg 2001;20:961–6.

[8] Yap CH, Reid C, Yii M, Rowland MA, Mohajeri M, Skillington PD, Seevanayagam S, Smith JA. Validation of the EuroSCORE model in Australia. Eur J Cardiothorac Surg 2006;29:441–6.

[9] Chen C-C, Wang C-C, Hsieh S-R, Tsai H-W, Wei H-J, Chang Y. Application of European system for cardiac operative risk evaluation (EuroSCORE) in coronary artery bypass surgery for Taiwanese. Interact Cardiovasc Thorac Surg 2004;3:562–5.

[10] Jin R, Grunkemeier GL, Providence Health System Cardiovascular Study Group. Does the logistic EuroSCORE offer an advantage over the additive model? Interact Cardiovasc Thorac Surg 2006;5:15–7.

[11] Bhatti F, Grayson AD, Grotte G, Fabri BM, Au J, Jones M, Bridgewater B, North West Quality Improvement Programme in Cardiac Interventions.

The logistic EuroSCORE in cardiac surgery: how well does it predict operative risk? Heart 2006;92:1817–20.
[12] Shanmugam G, west M, Berg G. Additive and logistic EuroSCORE performance in high risk patients. Interact Cardiovasc Thorac Surg 2005;4:299–303.
[13] Zingone B, Pappalardo A, Dreas L, Logistic versus additive EuroSCORE. A comparative assessment of the two models in an independent population sample. Eur J Cardiothorac Surg 2004;26:1134–40.
[14] Ivanov J, Tu JV, Naylor CD. Ready-made, recalibrated, or remodeled? Issues in the use of risk indexes for assessing mortality after coronary artery bypass graft surgery. Circulation 1999;99:2098–104.
[15] D'Errigo P, Seccareccia F, Rosato S, Manno V, Badoni G, Fusco D, Peducci CA, the Research Group of the Italian CABG Outcome Project. Comparison between an empirically derived model and he EuroSCORE system in the evaluation of hospital performance: the example of the Italian CABG Outcome Project. Eur J Cardiothorac Surg 2008;33:325–33.

Editorial comment

Predicting operative risk: a worthy task — an elusive goal

Keywords: Outcomes; Mortality; Cardiac surgery

Risk scoring models are tools that have been developed to anticipate expected patient mortality with certain surgical procedures. These scores can be used in many ways: to compare operative results from different centres, as a tool for performance improvement within individual institutions or to select appropriate patients for inclusion or exclusion in clinical trials [1]. Fundamental to this process, however, is the accuracy with which the model performs in predicting mortality. The European System for Cardiac Operative Risk Evaluation (EuroSCORE) is one such instrument designed to predict operative mortality, and subsequently enable the reporting of risk-adjusted mortality as opposed to raw mortality. This tool was developed using patient demographics and operative outcomes in 19,030 patients from 128 centres performing cardiac surgery [2]. The additive EuroSCORE was the first model available for use. The logistic model was subsequently developed to provide greater precision in higher-risk patients. While several reports validate its findings in patients undergoing coronary artery bypass grafting, increasing concerns over its apparent overestimation of mortality in higher-risk patients has led to a re-evaluation of its effectiveness.

The article by Ranucci and colleagues in this journal is an attempt to verify the accuracy and precision of the logistic EuroSCORE (LogES), as well as create a 're-calibration' adjustment to increase its accuracy [3]. They acknowledge that estimation of mortality in higher-risk patients is poor using the LogES, and graphically demonstrate multiple areas of divergence of observed mortality from expected within a population of patients with varying risk factors and procedures. A LogES of 5 was chosen as the representative point of divergence to indicate a high-risk patient, although clinically, most practitioners would not assume that these patients would be at increased risk. In fact, in a recent report by Leontyev and colleagues looking at risk stratification for aortic valve replacement in octogenarians, a logistic EuroSCORE of <10 was considered low risk [4].

The authors create a re-calibration factor in order to improve the accuracy of the LogES by drawing linear regression lines through various points, and subsequently creating a 'fixed' correction factor for each range of predicted mortality. This correction is then used to re-calculate an adjusted LogES using various multipliers depending on the range of predicted mortality. The validity of this type of model manipulation comes into question, however, by observing that in the new model the splines do not meet. Logistic regression equations should not have the variability or 'bumps' that are observed in the new adjusted LogES. It is difficult to 'reverse engineer' the risk algorithm based on observed mortality in a localised population because the problem lies with the original model itself. Clearly, the LogES diverges from observed mortality for higher-risk patients by over-estimating mortality, sometimes greatly. However, it is probably just as inaccurate at the low end of risk as well, but the differences are less obvious since the values tend to clump closer to zero. The EuroSCORE was predominately derived from a population of patients undergoing coronary artery bypass grafting (CABG). As a consequence, the model is designed such that CABG is considered the baseline risk, and any other procedure such as valve replacement or repair becomes 'procedure other than CABG'. Consequently, mitral valve repair or replacement, aortic valve replacement or any valve procedure combined with CABG will result in the same predicted risk despite obvious differences in the complexity of the procedures [5]. Given the baseline risk of CABG inherent in the model, it becomes clear why relatively lower-risk procedures such as isolated aortic valve replacement can demonstrate elevated predicted risk. Furthermore, this poor calibration of the EuroSCORE has been found to progressively worsen with era of operation [6] Given that the EuroSCORE is based on data collected in 1995, it would be surprising for it to maintain calibration to contemporary results. Numerous improvements and refinements in surgical technique and perioperative care have led to a reduction in operative mortality not reflected in the data set used to develop the EuroSCORE. A better approach than trying to create a re-calibration factor based on old data would be to update the data set and