

Validation of the EuroSCORE Probabilistic Model in Patients Undergoing Coronary Bypass Grafting

Sarah Lafuente,^a Antoni Trilla,^a Laia Bruni,^a Raquel González,^a María J. Bertrán,^a José Luis Pomar,^b and Miguel A. Asenjo^a

^aServicio de Medicina Preventiva y Epidemiología, Hospital Clínic, IDIBAPS, Universidad de Barcelona, Barcelona, Spain

^bServicio de Cirugía Cardíaca y Vascul ar, Hospital Clínic, Universidad de Barcelona, Barcelona, Spain

Introduction and objectives. EuroSCORE utilizes a probabilistic model for predicting the risk of in-hospital mortality in patients undergoing cardiac surgery. It is a useful instrument for evaluating quality of care. The model has two variants: the logistic EuroSCORE and the additive EuroSCORE. The aim of this study was to validate the EuroSCORE model in patients undergoing surgery at Hospital Clínic in Barcelona, Spain, and to compare the results obtained with the 2 variants.

Methods. The study included all patients who received a coronary artery bypass graft (CABG) at Hospital Clínic in Barcelona in 2 consecutive years. The model's validity was assessed on the basis of its calibration (using the Hosmer-Lemeshow test) and its discrimination (using the receiver operating characteristic [ROC] curve). The 2 models were compared by carrying out a descriptive analysis of mortality for the whole group and for different risk groups, and by determining the models' discriminative power.

Results. A total of 498 patients underwent CABG surgery and were included in the study. The Hosmer-Lemeshow test showed that the model's calibration was satisfactory ($P=0.32$) and the area under the ROC curve was 0.83. The observed in-hospital mortality rate was 5.8%. The predicted rate was 4.2% with the logistic EuroSCORE and 3.9% with the additive EuroSCORE. Large differences were observed in high-risk patients. In these patients, the mortality predicted by the logistic variant was closer to the actual mortality.

Conclusions. EuroSCORE's validity was found to be satisfactory and the model can be used to evaluate quality of care. In high-risk patients, mortality estimated using the logistic model was closer to the actual mortality.

Key words: Probabilistic model. Quality evaluation. In-hospital mortality. Coronary artery bypass grafting.

SEE EDITORIAL ON PAGES 567-71

Correspondence: Dr. A. Trilla.
Servicio de Medicina Preventiva y Epidemiología. Hospital Clínic.
Villarroel, 170. 08036 Barcelona. España.
E-mail: atrilla@clinic.ub.es

Received August 11, 2007.

Accepted for publication December 19, 2007.

Validación del modelo probabilístico EuroSCORE en pacientes intervenidos de injerto coronario

Introducción y objetivos. EuroSCORE es un modelo probabilístico para estimar la probabilidad de mortalidad hospitalaria en pacientes sometidos a cirugía cardíaca. Es un instrumento útil para evaluar la calidad asistencial. Existen dos variantes del modelo, el EuroSCORE logístico (EU-L) y el EuroSCORE aditivo (EU-A). El objetivo del estudio es validar el modelo EuroSCORE en pacientes intervenidos en el Hospital Clínic de Barcelona y comparar los resultados de las dos variantes del modelo.

Métodos. Se ha incluido a los pacientes intervenidos de injerto coronario en el Hospital Clínic de Barcelona durante 2 años consecutivos. Se ha validado el modelo a partir de su capacidad de calibración (prueba de Hosmer-Lemeshow) y discriminación (área bajo la curva ROC). Se han comparado los dos modelos con un análisis descriptivo de la media de la mortalidad para el total y según grupos de riesgo y mediante su poder de discriminación.

Resultados. Un total de 498 pacientes fueron intervenidos e incluidos en el estudio. La calibración del modelo es satisfactoria ($p = 0,32$) y el área bajo la curva ROC es de 0,83. La mortalidad hospitalaria observada alcanzó el 5,8% y la estimada, el 4,2% (EU-L) y el 3,9% (EU-A). Se han observado mayores diferencias en el grupo de pacientes de alto riesgo, en los que la mortalidad predicha por la variante logística se aproxima más a la mortalidad real.

Conclusiones. EuroSCORE ha sido validado adecuadamente y puede utilizarse para medir los resultados de la práctica asistencial. El modelo logístico se aproxima más a la mortalidad real en el grupo de pacientes de alto riesgo.

Palabras clave: Modelo probabilístico. Evaluación de resultados. Mortalidad hospitalaria. Injerto coronario.

ABBREVIATIONS

HCB: Hospital Clínic of Barcelona
 EU-A: additive EuroSCORE
 EU-L: logistic EuroSCORE
 EuroSCORE: European System for Cardiac
 Operative Risk Evaluation
 ROC: receiver operating characteristic

INTRODUCTION

Health-care providers need to be able to reliably assess their activities in terms of outcomes, quality, and cost-effectiveness. This is, in part, due to the constant rise in health care costs and the fact that resources are limited, though other drivers include the increased demand for health services and the need to compare clinical outcomes between centers. As all institutions require information on quality of care they need to be able to summarize their activities in terms of outcomes adjusted by the center's specific characteristics.¹

In the area of heart surgery, both in Europe² and the United States,³ the periodic publication of outcomes reports is now widespread and adequately regulated.^{4,5} These reports incorporate information from mathematical models which are used to predict the likelihood of a certain event, such as death, occurring in a given individual based on a group of risk factors attributed to that particular patient.⁶

The European System for Cardiac Operative Risk Evaluation (EuroSCORE) is a logistic model which is used to predict hospital mortality in patients undergoing a cardiac intervention. Using 18 risk variables and a beta coefficient associated with each variable (Table 1), the model provides the likelihood of death for any individual. The model was created and initially validated in a cross-sectional study^{7,8} of 19 030 European patients in 1999. Since then, it has become the most widely used model worldwide in this type of patient.

A much simpler variant of the logistic model is the additive EuroSCORE, which assigns a weight to each risk factor presented by the patient. The sum of the weights provides the likelihood of dying for that patient. The widespread and uniform use of a single probabilistic model allows for internal and external comparisons over time and can help to minimize risk adverse behavior which might be fomented if comparisons are made using unadjusted outcomes.^{9,10}

The primary objective of the present study was to validate this predictive model of mortality in a large teaching hospital in Barcelona, Spain. The model was assessed in terms of fit and discriminatory capacity.¹¹

TABLE 1. EuroSCORE Risk Factors With Corresponding Weights and Beta Coefficients

Variable (%)	Additive Weight	β
Age	1 for every 5 years >60	.0666354
Female	1	.3304052
Creatinine >200 μ mol/L	2	.6521653
Extracardiac arteriopathy	2	.6558917
COPD	1	.4931341
Neurological dysfunction	2	.841626
Previous cardiac surgery	3	1.002625
Active endocarditis	3	1.101265
Critical preoperative state	3	.9058132
Unstable angina	2	.5677075
LVEF <30%	3	1.094443
LVEF 30%-50%	1	.419643
Recent myocardial infarct	2	.5460218
Pulmonary hypertension PA >60 mm Hg	2	.7676924
Emergency operation	2	.7127953
Post-infarct septal rupture	4	1.462009
Other than isolated CABG	2	.5420364
Surgery on thoracic aorta	3	1.159787

COPD indicates chronic obstructive pulmonary disease; LVEF, left ventricular ejection fraction; PA, pulmonary artery.

A second study objective was to compare the additive and logistic versions of the model and to determine which was the most appropriate for use in different groups of patients defined by their level of risk. The additive version has been the most widely used of the model variants because, although it is less precise, it is much easier to calculate and it can be calculated at the bedside. Nevertheless, the logistic equation has been shown to better predict mortality, particularly in high risk patients, and is recommended for use in those patients.¹²⁻¹⁴

METHODS

Patients

This was a validation study performed in the Hospital Clínic of Barcelona (HCB) using a retrospective design. The hospital is a 720-740 bed center which deals with 40 000 admissions annually. Patients with cardiovascular disease are attended in the Clinical Institute for Diseases of the Thorax which includes, among others, the Cardiovascular Surgery and Cardiology services. The study was approved by the Institutional Review Board of the HCB.

The center's computerized database (SAP[®]) was used to obtain data on all patients who underwent procedures defined in the CIE-9-CM¹⁵ and who were assigned codes 36.10 to 36.17 and 36.19. The SAP is used for both clinical and administrative purposes and has high reliability. All discharges are coded. The study population

consisted of patients derived for coronary surgery and who underwent surgery between July 1, 2004 and July 1, 2006.

Variables and Statistical Analysis

The 18 variables included in the EuroSCORE predictive model were identified for all patients included in the study (Table 1) together with administrative variables (date of admission and discharge) and data on deaths occurring while patients were hospitalized in relation to the intervention. These data were obtained from computerized clinical records which included discharge history as well as surgery, pre-anesthesia, and laboratory reports. If no value was recorded for a specific risk factor, it was assumed that the risk factor was absent.

Both the logistic and additive versions of the EuroSCORE¹² were used to predict mortality. In the case of the additive EuroSCORE, the probability of dying was calculated by summing the relative weights for each risk factor for all individuals.

In order to calculate predicted mortality using the logistic model, the following equation was used:

$$\text{Mortality} = e^{(\beta_0 + \sum \beta_i \chi_i)} / (1 + e^{(\beta_0 + \sum \beta_i \chi_i)})$$

where β_0 takes a value of -4.789594 (logistic regression constant) and β_i is the regression coefficient for the variable χ_i in Table 1. For the age variable, in the logistic method β was multiplied by the number of years that the patient exceeded 60 years of age. In the additive method, a weight of 1 was assigned for every 5 years (or part of 5 years) over 60.

The validity of the logistic regression model was analyzed by examining its goodness of fit and discriminatory capacity. Goodness of fit was assessed using the Hosmer-Lemeshow test¹⁶⁻¹⁹ which estimates a C statistic from the difference between observed and expected values for mortality in different risk groups. The lower the C statistic the better the model's fit. A value of $P > .05$ indicates that the model fits the data well and that it therefore accurately predicts mortality. The test is most frequently used to validate recently created models but it is equally useful in validating an existing model which has been applied in a new set of data, as in the present study.

The model's ability to discriminate is assessed in terms of its capacity to distinguish between patients who died during hospitalization from those who did not. Discriminatory capacity was analyzed by calculating the area under the ROC curve. A value of 0.5 indicates that the model is equivalent to pure chance and a value of 1 indicates perfect discrimination.¹⁷

The logistic and additive models were compared by calculating the mortality predicted by each in both the overall sample and in 2 sub-groups defined by level of risk. The high and low risk groups were defined using a

cut-off point on the additive EuroSCORE of 6 points,⁶ a cut point which had previously been used for this purpose.^{8,12} ROC curves were calculated for both models.

The validation analysis was carried out using the STATA[®] v.8 statistical package and the comparative analysis using SPSS[®] v.12.0.

RESULTS

A total of 498 patients underwent heart surgery in HCB between July 1, 2004 and June 30, 2006. The distribution of EuroSCORE risk variables in these patients is summarized in Table 2.

For the logistic model, a C statistic of 11.51 ($P = .32$) was obtained on the Hosmer-Lemeshow test and the area under the ROC curve was 0.83 (Figure 1).

Among the patients who underwent surgery during the study period, there were 29 hospital deaths, giving an overall mortality rate of 5.8%. Total predicted mortality was 3.9% using the additive model and 4.2% using the logistic model (Table 3).

In the low risk group ($n = 412$), observed mortality was very similar to that predicted by both models; however, in the high risk group ($n = 86$) the values predicted by the logistic model were closer to observed values (Table 3). Both models showed good discriminatory power, with an area under the ROC curve of 0.84 for the additive model and 0.83 for the logistic model (Figures 1 and 2).

DISCUSSION

In order to evaluate the quality of health care services and to be able to adequately inform patients on the likely outcomes of the health care process, crude values for overall observed or expected outcomes are often not sufficient. Prognostic models which take into account patients' specific characteristics and which provide risk-adjusted outcomes for interventions are required and more useful.⁶

There are many risk-adjusted models available to predict mortality in cardiac surgery interventions, though in recent years the EuroSCORE⁸ has become one of the most widely used in western countries.

The HCB is a reference center for this type of heart surgery and performs a large number of coronary interventions. In this type of center, reliable and comparable data are required in order to assess the quality of care.

Before using a probabilistic model in a context other than that for which it was created,²⁰ the model should be validated to ensure that it does not generate erroneous probabilities. The aim of the present study was to validate²¹ the EuroSCORE in the HCB.

The C statistic obtained with the Hosmer-Lemeshow was $P = .32$, which indicates satisfactory model fit for patients undergoing surgery in our center. The model's discriminatory power was also adequate, as indicated by

TABLE 2. Risk Variables in the EuroSCORE Model; Prevalence in Present Study and in Initial Model Development

Variable	Patients (n=498), No.	Percentage	EuroSCORE Data-Base (19030), %
Age			
<60 y	150	30.1	33.2
60-64 y	68	13.7	17.8
65-69 y	74	14.9	20.7
70-74 y	115	23.1	17.9
>74 y	91	18.11	9.6
Female	71	14.3	27.8
Creatinine >200 µmol/L	18	3.6	1.8
Extracardiac arteriopathy	61	12.2	11.3
COPD	44	8.8	3.9
Neurological dysfunction	10	2	1.4
Previous cardiac surgery	5	1	7.3
Active endocarditis	4	0.8	1.1
Critical preoperative state	12	2.4	4.1
Unstable angina	7	1.4	8
LVEF <30%	21	4.2	5.8
LVEF 30%-50%	89	17.9	25.6
Recent myocardial infarction	113	22.7	9.7
Pulmonary hypertension PA>60 mm Hg	0	0	2
Emergency operation	13	2.6	4.9
Post-infarct septal rupture	0	0	0.2
Other than isolated CABG	100	20.1	36.4
Surgery on thoracic aorta	4	0.8	2.4

COPD indicates chronic obstructive pulmonary disease; LVEF, left ventricular ejection fraction; PA, pulmonary artery.

TABLE 3. Observed and Predicted Mortality (95% CI) for the Additive and Logistic EuroSCORE (%), Overall Sample, and by Risk Group^a

	Patients, No.	Observed	Additive EuroSCORE	Logistic EuroSCORE
Total	498	5.82 (3.76-7.89)	3.88 (3.64-4.11)	4.24 (3.79-4.69)
Low risk (ES≤6)	412	3.15 (1.46-4.85)	3 (2.81-3.19)	2.74 (2.51-2.97)
High risk (ES>6)	86	18.6 (10.21-27)	8.08 (7.76-8.41)	11.41 (9.70-13.13)

^aES indicates additive EuroSCORE.

an area under the ROC curve of 0.83. Given these results, we can conclude that the EuroSCORE model has been validated for use in this center and that it has proven to be a reliable instrument. This signifies that the model's predictions of the probability of dying are valid and appropriately risk-adjusted for surgery patients in HCB. It should be pointed out that mortality during hospitalization is a very favorable measure of mortality as it does not incorporate mortality after discharge. Nevertheless, we believe that intervention-related mortality after discharge was practically null in this series.

The results of this study are in line with those reported in the earlier article on model validation in six countries in the European Union,²² to which Spain contributed 2422 patients. In the earlier study, the Hosmer-Lemeshow C statistic was $P=.38$ and the area under the ROC curve was 0.87. Nevertheless, the earlier validation in Spain

took place in very diverse settings and conditions, which meant there was a need for further validation in specific contexts before the model could be used with confidence.

The additive, or standard, model has been the most widely used because it is easy to use. It is a simplified version of the logistic model and the weights it uses are derived from that model. A previous study¹² indicated that the additive model tended to underestimate the probability of death in high risk patients.

In the comparison of the 2 models, only low and high risk groups were studied because of the relatively low mortality in the study population. Creating a larger number of risk groups would have led to very broad confidence intervals for the predicted mortality rates and would have hindered comparisons. The cut point for defining the 2 groups was a EuroSCORE value on the additive model of >6. In this study, we found that the logistic model

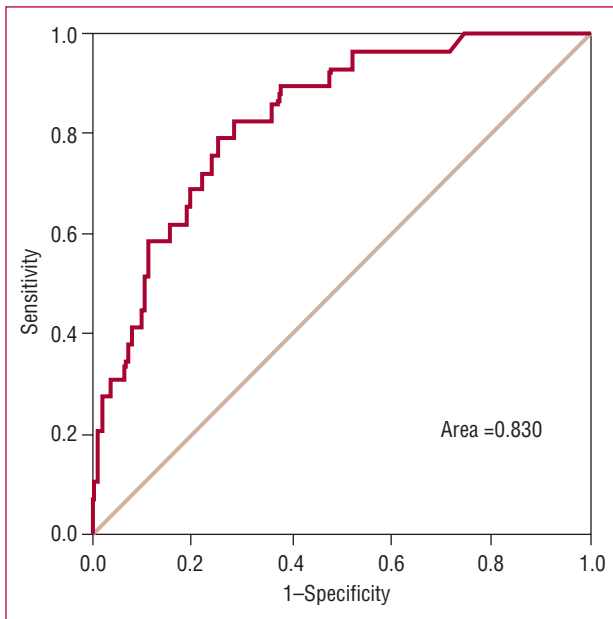


Figure 1. ROC curve for EuroSCORE logistic model.

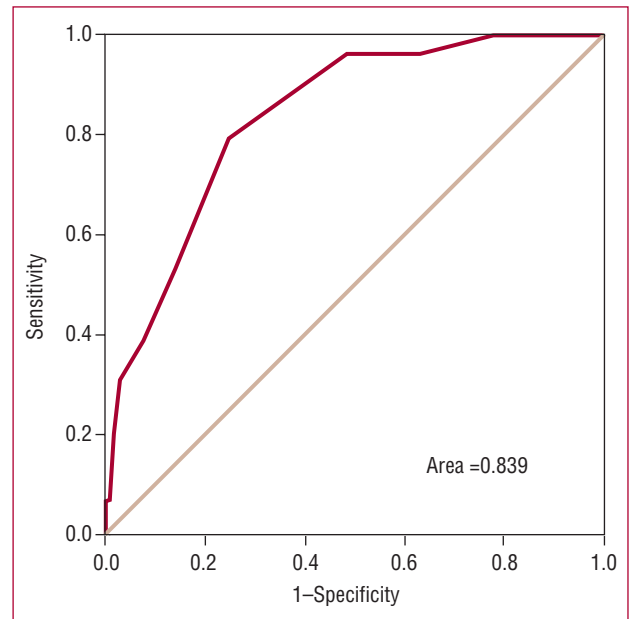


Figure 2. ROC curve for EuroSCORE additive model.

more accurately predicted the probability of death in the high risk group, a result in line with previous studies which have compared the 2 models.¹²

These results indicate that the logistic model is more accurate and more appropriate for use in daily practice in heart surgery units. The fact that most such units now have the technical resources to rapidly calculate a score using the logistic model further supports its use.

A study in a larger sample would generate a larger number of events (deaths) and would provide more solid results as the Hosmer-Lemeshow test is based on contrasting predicted events with observed events.

A direct comparison of real and expected events was not carried out as they are two distinct variables which provide information about events. True mortality describes the event death, whether observed or otherwise, for each patient (dichotomous variable). Predicted mortality, on the other hand, indicates the likelihood of dying for each patient based on specific characteristics included in the model (quantitative continuous variable). The analysis of the relationship between these 2 variables is precisely what constitutes the validation of the model. The validated model is useful because it allows us to perform risk assessments for patients which can then be compared with observed outcomes, while taking into account the level of risk.

The publication of the outcomes of health care has been a reality in the United States and the United Kingdom for over 15 years. The impact of such publications has been extensively analyzed, with the results showing a clear improvement in the quality of health care.²³ The instrument validated in the present study could be useful in providing systematic information on the outcome of

interventions in centers providing the relevant services. The information could be published and made available both to citizens and the purchasers of health care.

CONCLUSIONS

The results of this study allow us to conclude that the EuroSCORE is a useful probabilistic model in this public, teaching hospital. It can be used to estimate the probability of death in patients scheduled for heart surgery and to assess the outcomes of health care. The logistic model is the most reliable of the 2 versions, particularly in high risk patients.

REFERENCES

1. Chassin MR, Hannan EL, DeBuono BA. Benefits and hazards of reporting medical outcomes publicly. *N Engl J Med.* 1996;334:394-8.
2. Rowan K, Harrison D, Brady A, Black N. Hospitals' star ratings and clinical outcomes: ecological study. *BMJ.* 2004;328:924-5.
3. Davies HT, Marshall MN. Public disclosure of performance data. *Lancet.* 1999;353:1639-40.
4. Keogh B, Spiegelhalter D, Bailey A, Roxburgh J, Magee P, Hilton C. The legacy of Bristol: public disclosure of individual surgeons' results. *BMJ.* 2004;329:450-4.
5. Keogh BE, Dussek J, Watson D, Magee P. Public confidence and cardiac surgical outcomes. *BMJ.* 1998;316:1759-60.
6. Asimakopoulos G, Al-Ruzzeah S, Ambler G, Omar RZ, Punjabi P, Amrani M, et al. An evaluation of existing risk stratification models as a tool for comparison of surgical performances for coronary artery bypass grafting between institutions. *Eur J Cardiothorac Surg.* 2003;23:935-41.
7. Roques F, Nashef SA, Michel P, Gauducheau E, de Vincentiis C, Baudet E, et al. 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.
8. 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;6:9-13.
 9. Bridgewater B, Grayson AD, Au J, Hassan R, Dihmis WC, Munsch C, et al. Improving mortality of coronary surgery over first four years of independent practice: retrospective examination of prospectively collected data from 15 surgeons. *BMJ.* 2004;329:421.
 10. Treasure T. Lessons from the Bristol case. *BMJ.* 1998;163:1685-6.
 11. Wade A. Derivation versus validation. *Arch Dis Child.* 2000;83:459-60.
 12. 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.
 13. Bridgewater B, Grayson AD, Jackson M, Brooks N, Grotte GJ, Keenan J, et al. North West Quality Improvement Programme in Cardiac Interventions. Surgeon specific mortality in adult cardiac surgery: comparison between crude and risk stratified data. *BMJ.* 2003;327:13-7.
 14. Sergeant P, de Worm E, Meyns B. Single centre, single domain validation of the EuroSCORE on a consecutive sample of primary and repeat CABG. *Eur J Cardiothorac Surg.* 2001;20:1176-82.
 15. The International Classification of Diseases, 9th Revision, Clinical Modification (ICD-9-CM), Sixth Edition. Free online searchable 2004 ICD-9-CM and Medical Terminology Dictionary [citado 28 Ago 2005]. Disponible en: <http://icd9cm.chrisendres.com/index.php>
 16. Lemeshow S, Hosmer D. A review of goodness of fit statistic for use in the development of logistic regression models. *Am J Epidemiol.* 1982;92:106.
 17. Hosmer DW, Lemeshow S. *Applied logistic regression.* 2nd ed. New York: John Wiley & Sons; 2000.
 18. Rué M. Apunts metodològics sobre els models probabilistics. *Annals de Medicina.* 2004;87:10-1.
 19. Lemeshow S, Klar J, Teres D. Outcome prediction for individual intensive care patients: useful, misused, or abused? *Intensive Care Med.* 1995;21:770-6.
 20. Bhatti F, Grayson AD, Grotte G, Fabri BM, Au J, Jones MT, et al. The logistic EuroSCORE in cardiac surgery: how well does it predict operative risk? *Heart.* 2006;92:1817-20.
 21. Hosmer DW, Taber S, Lemeshow S. The importance of assessing the fit of logistic regression models: a case study. *Am J Public Health.* 1991;81:1630-5.
 22. 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.
 23. Hannan EL, Kilburn H, Racz M, Shields E, Chassin MR. Improving the outcomes of coronary artery bypass grafting surgery in New York State. *JAMA.* 1994;271:761-6.