

## Criteria for Using Risk Models in Cardiac Surgery

José M. Cortina Romero

Servicio de Cirugía Cardíaca, Hospital 12 de Octubre, Madrid, Spain

Few therapeutic interventions have been so widely researched, evaluated, and commented on as cardiac surgery, and in particular coronary revascularization. Aspects related to outcomes and the appropriateness of indications have, among many others, received much attention.

This type of research is driven by many interest groups, including heart surgeons, cardiologists, epidemiologists, health care managers, health service providers, and the media. It is one of the few specialties in which outcome audits are mooted as soon as the alarm bells sound.

Why has heart surgery been the object of such exhaustive analysis compared to other interventions? Although it is not the aim of this editorial to go into detail on this point, one of the reasons might be the interest shown by heart surgeons themselves in evaluating their own activities.

Groups performing cardiac surgery have pioneered the introduction of predictive systems to estimate the risks associated with their surgical procedures. This has led to a culture centered around “risk” and risk-assessment and it is nowadays common for heart surgeons and cardiologists to factor risk assessments into decision-making with their surgery patients. This is often done by employing risk assessment systems such as the EuroSCORE, Parsonnet, or STS.<sup>1-5</sup> The 2004 AHA/ACC clinical guidelines<sup>6</sup> suggest that preoperative predictive systems to assess risk can help both clinicians and patients to understand the risk-benefit equation for the procedure concerned (class IIa recommendation, level of evidence C).

SEE ARTICLE ON PAGES 589-94

Correspondence: Dr. J.M. Cortina Romero.  
Servicio de Cirugía Cardíaca. Hospital 12 de Octubre.  
Avda. Córdoba, s/n. 28041 Madrid. España.  
E-mail: jcortina.hdoc@salud.madrid.org

Apparently, no equivalent recommendation exists for other procedures such as coronary revascularization using endoluminal procedures. The lack of such a recommendation cannot be ascribed, however, to an absence of extensively validated, predictive models for percutaneous coronary procedures.<sup>7,8</sup> I believe it would be desirable for such an approach to be extended to other therapeutic procedures, and not only in the area of heart disease. Nevertheless, it is clear that the culture required for the application and generalization of this methodology is lacking.

It might be pretentious to try to establish, here, recommendations for the use of this type of tool. But the importance of the conclusions derived from their application suggests that we need to follow strict criteria when applying these tools in order to guarantee the validity of our conclusions. If strict criteria are not followed, erroneous interpretations may result which, instead of helping to clarify things, will lead to greater confusion. The conditions for applying these tools should be even stricter when they are used in quality assessment.

Predictive models are used at 3 different levels: individual level estimations, estimations in specific disease groups, and estimations in complete series. The latter reflect the activity of a given surgical department or unit.

Although use at the first 2 levels is acceptable, and we'll comment on these later, these statistical tools were principally developed to monitor activity in complete series.

The final aim in developing this type of model has always been to provide a tool which would allow objective evaluations and comparisons of outcomes and case-mix in populations of heart surgery patients. Instruments such as these are required as part of the philosophy of health care quality assessment, particularly when the current hypothesis regarding the assessment of care infers that if we can ensure comparable populations, any significant differences in outcome will be due to differences in the quality of care. An objection to this point of view could be that structural differences between groups are not taken into account. Unfortunately, at present there are no models available which adjust for this important dimension.

The publication of the article entitled “Validation of the EuroSCORE Probabilistic Model in Patients

Undergoing Coronary Bypass Grafting”<sup>9</sup> raises several issues related to the application of this type of tool. As mentioned above, the use of scores or predictive models has become routine in many heart surgery departments. Nevertheless, there may still be sufficient confusion surrounding their use to consider it is not yet sufficiently standardized. In the hope of clarifying some issues related with the methodology, I would like to comment on the following aspects:

- The populations in which the models are applied
- The issue of whether the logistic or additive EuroSCORE is more appropriate
- Model application in administrative databases compared to clinical databases
- What it means to validate these models
- Interpreting the results of applying the models

Although the following comments could apply to any of the predictive tools currently available, we will generally refer to the EuroSCORE because it is the most widely used in Spain.

### Populations in Which Models Are Applied

Currently, these systems are being used to estimate individual risk, to estimate risk in a given disease group, and to estimate risk in complete series of surgical patients.

These models are usually constructed using logistic regression techniques to identify factors with the greatest impact on the dependent variable, in this case hospital mortality. Although multivariable analysis applied in large populations can identify many factors associated with hospital mortality, when constructing the models only a limited number of factors are generally used. There are several reasons for this, including the fact that, beyond a certain point, goodness of fit and discriminatory capacity are not usually greatly improved by the addition of more factors. Models also generally need to be simple. Other reasons are purely methodological. Harrel et al<sup>10</sup> concluded that, in models based on logistic regression, there should be at least 10 events (deaths) per prognostic variable in the model. As the “event” to be predicted in these models is usually death, it is difficult to always meet this criterion, even in large populations.

As regards their use for individual level estimations of risk, given the limited number of variables in the EuroSCORE model, it can only be considered a rough guide at this level. This is particularly true if the patient presents a clear risk factor for death which is not included in the model. Nevertheless, for the majority of patients scheduled for heart surgery, such as those with isolated coronary or valve disease, or a combination of these, the logistic EuroSCORE is useful in identifying the likelihood of risk in a given case. At individual level, it would be

desirable to have a model which takes many variables into account; however, the only model which currently incorporates a high number of variables is the STS model. It should be remembered that indicating or rejecting surgery based on estimates of high risk at individual level may not be appropriate.

Applying such models in groups of specific diseases, such as coronary disease, valve disease, etc, leads to another problem, which is different from that arising at individual level. Studies in several series have concluded that specific models are required for each disease group, particularly in the case of valve surgery patients.<sup>11-14</sup> Of course, the ideal situation would no doubt be to have one model for each type of patient, but these specific models will need to be appropriately constructed.

It should also be made very clear in which population the model is being applied. The article by Lafuente et al does not make this clear. The title suggests that the population to be studied will be isolated coronary artery bypass patients, but then there are 4 patients with endocarditis and in 100 patients other surgical procedures were involved.

In principal, the most appropriate application for the EuroSCORE and the one for which it was created is the study of complete series of surgical patients. The basic objective in this case is to provide a standardized tool for quality control. The tool should be applied in any patient who underwent cardiac surgery using extracorporeal circulation. The only other group in which its use seems reasonable are patients who have received coronary surgery without extracorporeal circulation, as the tool was been extensively validated in this population.<sup>15,16</sup> When used in this way, several conditions need to be strictly adhered to: all patients in which the tool can be applied should be included, all data relating to factors included in the model should be collected, and, above all, all related deaths should be recorded.

### Which is the Appropriate Model: The Logistic or Additive EuroSCORE?

As mentioned previously, these tools are basically designed for use in quality assessment,<sup>17</sup> which means they should be applicable to all patients. For this purpose, and although there are many models available, at present I believe that the EuroSCORE should continue to be used. There are many reasons for this including the fact that it has been extensively validated in different settings,<sup>18,19</sup> the fact that Spain provided 10% of the population from which the model was developed, the fact that it is simple to use, and that it can be used in the usual population of cardiac surgery patients. The authors of the original instrument are currently planning to develop a new version of the EuroSCORE using the same methodology applied in the original study.

With respect to using the EuroSCORE, there is some confusion regarding whether the logistic or additive model

should be used. It should be remembered that both models are essentially the same and that the additive model is a simplification of the logistic model, not a variant of it. The additive model assimilates and simplifies the beta coefficients used in the logistic regression equation so that it is simpler to use. Using a simple sum approach to estimate probabilities may nevertheless be a source of error, particularly in high risk patients. In order to assign a probability of death to a given score, Table should be used.<sup>1</sup>

A comparison of the additive and logistic versions of the EuroSCORE has already been performed,<sup>20</sup> though the results generated some confusion and the conclusions derived can be easily understood by observing the Table. Leaving aside the issue of model calibration in high risk groups, which, incidentally, usually only form a small part of the overall experience of a surgical group, the discriminatory power of the 2 versions is identical.

In summary, the logistic model should ideally be used in order to avoid confusion. If the additive model is used, particularly for individual level estimates of risk, the risk of dying associated with each score segment should be taken into account.

### Use With Administrative or Clinical Databases

Another interesting aspect is the use of predictive models with administrative compared to clinical databases. The majority of existing models have been developed using the latter, though they are increasingly used with administrative databases. At present, these databases do not allow for a rigorous application of predictive tools. This means that any conclusions drawn from such studies should be treated with caution.

Although the managers of administrative databases frequently claim that they are highly reliable, the reality is that, at present, they may not be sufficiently reliable. There are many reasons for this, notably:

- Administrative databases usually record data retrospectively. This implies relatively high levels of missing data which in turn may require potentially

unacceptable assumptions to be made.<sup>21,22</sup> Prospective data collection is clearly the ideal here

- Patients are often wrongly classified which means it is difficult to adequately monitor the pathology in question<sup>23,24</sup>

- The lack of standardized end-points is another source of error.<sup>25,26</sup> The standard definition of hospital mortality should be applied, ie, mortality occurring during hospitalization or in the first 30 days after the intervention. It is not acceptable to state that mortality outside of the hospital stay was minimal based only on subjective beliefs

Obviously, it would be ideal if there were no discrepancies due to the type of database used. More desirable again would be the existence of a single information management tool whose users would apply standardized, shared criteria when exploiting the data.

Independently of the tool used, the value of any conclusions reached will depend on the external validity of the data and the data management system.

### What Does it Mean to Validate a Predictive Tool?

The word “validity” may well have been frequently misapplied when referring to this type of system. It is clearly wrong to claim to be validating a model simply because you use it in a particular set of circumstances. The validation of a probabilistic model requires compliance with a series of conditions which are only infrequently met during the application of such models.<sup>27</sup> From a statistical point of view, validating a prognostic model means showing that it functions correctly in a population other than that in which it was developed.

The first condition to be met when validating a model is to have an adequate sample size. Some authors<sup>28</sup> suggest that at least 100 events (deaths) are required. This implies that, if mortality is around 5%, a sample of 2000 patients will be needed.

As well as sample size, the score must be strictly applied to all patients, without assumptions having to be made because of missing data and, in particular, without missing data on the event which represents the dependent variable.

Finally, when a system has been so extensively validated as the EuroSCORE, an inability to validate the model may suggest 1 of 2 things. On the one hand, the tool may not have been correctly applied or, what is worse, the quality of care in the series studied is poor.

### Interpretation of Results

If the model used has been appropriately validated and correctly applied, the interpretation can only be related to the quality of care, with the single proviso mentioned earlier that the model does not take into account structural differences.

### Probability of Dying Based on the Additive EuroSCORE

Score	Mortality, %	95% CI
0, 1, 2	0.8	0.6-1.1
3, 4, 5	3	2.6-3.5
>5	11.2	10.2-12.2
Sub-groups >5		
6, 7, 8	7.4	6.5-8.4
9, 10, 11	16.4	14-19.1
>11	36.5	31-42.3

The only interpretation available if results are significantly above or below those expected is that the quality of care is either below or above par.

It is not only licit, but obligatory to carry out a statistical comparison of mortality observed in a given context with average observed mortality. These models were developed specifically with this in mind, and it is the way they are currently used.<sup>29</sup> Avoiding this type of comparison, as suggested by Lafuente et al in their article, is to veer from the essence of this type of methodology.

My belief is that scientific societies should establish precise guidelines regarding the use of these tools. I also believe that groups performing procedures which are susceptible to this type of analysis should incorporate predictive models into their daily practice. It is clear that inappropriate or incorrect use of these tools can lead to erroneous conclusions, an issue of particular sensitivity when quality assessment is involved.

## REFERENCES

- 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.
- Parsonnet V, Dean D, Bernstein AD. A method of uniform stratification of risk for evaluating the results of surgery in acquired adult heart disease. *Circulation* 1989;79:13-12.
- Bernstein AD, Parsonnet V. Bedside estimation of risk as an aid for decision-making in cardiac surgery. *Ann Thorac Surg.* 2000;69:823-8.
- Shroyer AL, Grover FL, Edwards FH. 1995 coronary artery bypass risk model: The Society of Thoracic Surgeons Adult Cardiac National Database. *Ann Thorac Surg.* 1998;65:879-84.
- Shroyer AL, Plomondon ME, Grover FL, Edwards FH. The 1996 coronary artery bypass risk model: the Society of Thoracic Surgeons Adult Cardiac National Database. *Ann Thorac Surg.* 1999;67:1205-8.
- Eagle KA, Guyton RA, Davidoff R, Edwards FH, Ewy GA, Gardner TJ, et al. ACC/AHA 2004 guideline update for coronary artery bypass graft surgery: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Committee to Update the 1999 Guidelines for Coronary Artery Bypass Graft Surgery). *Circulation* 2004;110:e340-437.
- Kunadian B, Dunning J, Das R, Roberts AP, Turley AJ, Morley R, et al. External validation of established risk adjustment models for procedural complications after percutaneous coronary intervention. *1. Heart.* 2007 Nov 21 [Epub ahead of print].
- Moscucci M, O'Connor GT, Ellis SG, Malenka DJ, Sievers J, Bates ER, et al. Validation of risk adjustment models for in-hospital percutaneous transluminal coronary angioplasty mortality on an independent data set. *J Am Coll Cardiol.* 1999;34:692-7.
- Lafuente S, Trilla A, Bruni L, González R, Bertrán MJ, Pomar JL, et al. Validación del modelo probabilístico EuroSCORE en pacientes intervenidos de injerto coronario. *Rev Esp Cardiol.* 2008;61:589-94.
- Harrell FE, Lee KL, Califf RM, Pryor DB, Rosati RA. Regression modelling strategies for improved prognostic modelling. *Stat Med.* 1984;3:143-52.
- Ambler G, Omar RZ, Royston P, Kinsman R, Keogh BE, Taylor KM. Generic, simple risk stratification model for heart valve surgery. *Circulation.* 2005;112:224-31.
- Grossi EA, Schwartz CF, Yu PJ, Jorde UP, Croke GA, Grau JB, et al. High-risk aortic valve replacement: are the outcomes as bad as predicted? *Ann Thorac Surg.* 2008;85:102-6.
- Jin R, Grunkemeier GL, Starr A. Validation and refinement of mortality risk models for heart valve surgery. *Ann Thorac Surg.* 2005;80:471-9.
- van Gameren M, Kappetein AP, Steyerberg EW, Venema AC, Berenschot EA, Hannan EL, et al. Do we need separate risk stratification models for hospital mortality after heart valve surgery? *Ann Thorac Surg.* 2008;85:921-30.
- Vázquez Roque FJ, Fernández TR, Pita S, Cuenca JJ, Herrera JM, Campos V, et al. Evaluación preoperatoria del riesgo en la cirugía coronaria sin circulación extracorpórea. *Rev Esp Cardiol.* 2005;58:1302-9.
- Wu Y, Grunkemeier GL, Handy JR Jr. Coronary artery bypass grafting: are risk models developed from on-pump surgery valid for off-pump surgery? *J Thorac Cardiovasc Surg.* 2004;127:174-8.
- Díaz de Tuesta I, Cuenca J, Fresneda PC, Calleja M, Llorens R, Aldámiz G, et al. No hay relación entre el volumen quirúrgico y la mortalidad en los servicios de cirugía cardíaca en España. *Rev Esp Cardiol.* 2008;61:276-82.
- Nashef SA, Roques F, Hammill BG, Peterson ED, Michel P, Grover FL, et al. Validation of European System for Cardiac Operative Risk Evaluation (EuroSCORE) in North American cardiac surgery. *Eur J Cardiothorac Surg.* 2002;22:101-5.
- Roques F, Nashef SA, Michel P, Pinna PP, David M, Baudet E, et al. Does EuroSCORE work in individual European countries? *Eur J Cardiothorac Surg.* 2000;18:27-30.
- Michel P, Roques F, Nashef SA. Logistic or additive EuroSCORE for high-risk patients? *Eur J Cardiothorac Surg.* 2003;23:684-7.
- Herbert MA, Prince SL, Williams JL, Magee MJ, Mack MJ. Are unaudited records from an outcomes registry database accurate? *Ann Thorac Surg.* 2004;77:1960-4.
- Mack MJ, Herbert M, Prince S, Dewey TM, Magee MJ, Edgerton JR. Does reporting of coronary artery bypass grafting from administrative databases accurately reflect actual clinical outcomes? *J Thorac Cardiovasc Surg.* 2005;129:1309-17.
- Glance LG, Dick AW, Osler TM, Mukamel DB. Accuracy of hospital report cards based on administrative data. *Health Serv Res.* 2006;41:1413-37.
- Shahian DM, Silverstein T, Lovett AF, Wolf RE, Normand SL. Comparison of clinical and administrative data sources for hospital coronary artery bypass graft surgery report cards. *Circulation.* 2007;115:1518-27.
- Hannan EL, Racz MJ, Jollis JG, Peterson ED. Using Medicare claims data to assess provider quality for CABG surgery: does it work well enough? *Health Serv Res.* 1997;31:659-78.
- Parker JP, Li Z, Damberg CL, Danielsen B, Carlisle DM. Administrative versus clinical data for coronary artery bypass graft surgery report cards: the view from California. *Med Care.* 2006;44:687-95.

27. Altman DG, Royston P. What do we mean by validating a prognostic model? *Stat Med.* 2000;19:453-73.
28. Harrel FE. *Regression models strategies.* New York: Springer-Verlag; 2001.
29. Ribera A, Ferreira-González I, Cascant P, Pons JM, Permanyer-Miralda G. Evaluación de la mortalidad hospitalaria ajustada al riesgo de la cirugía coronaria en la sanidad pública catalana. Influencia del tipo de gestión del centro (estudio ARCA). *Rev Esp Cardiol.* 2006;59:431-40.