

Importance of risk stratification models in cardiac surgery

Philippe Kolh

University Hospital, Liège, Belgium

This editorial refers to 'Comparison of 19 pre-operative risk stratification models in open-heart surgery'¹ by J. Nilsson *et al.*, on page 867

In a context of growing control of health-care expenditures, it is important to assess cardiac surgical results as precisely as possible. However, as patient population may differ significantly between institutions and geographic areas, comparison of absolute numbers, such as mortality rates, is inappropriate for cost-benefit analysis and comparison of results between institutions.^{1,2}

Therefore, various risk stratification models have been developed to correct for differences between populations and to allow comparison of actual outcome with predicted outcome.³ Those models are increasingly used to investigate patient outcomes in relation to pre-operative patient and disease characteristics. Such models estimate coefficients for each risk factor of mortality, which are translated to risk scores. Then, the scores assigned to each risk factor are added to calculate the overall risk score of mortality for a patient and to construct clinical risk groups. Reference to these groups can be made to adjust clinical decisions to individual patients, to compare surgical performances, and for patient counselling.

Health authorities, hospitals, medical practitioners, and patients are increasingly placing importance in those models, aimed at obtaining objective risk-adjusted prediction of mortality after cardiac surgery. Indeed, risk stratification models can detect and quantify differences and changes in the risk profiles of patients presented for cardiac surgery. Risk prediction allows a more objective assessment of the surgical indication in individual patients by facilitating accurate balancing of potential risks and benefits.⁴

Usually, such models do predict outcome more accurately in the original setting than when used for other patients populations. Indeed, there are significant differences with regard to the initial patient population on which the score design was based. The clinical database used for score development could have been derived from a single or several institutions and from one or several countries. Further differences include retrospective versus prospective data collection and whether a prospective validation study was performed after the score design. Rigorous validation can be obtained by evaluating the model on new data with patients undergoing surgery at a subsequent period and at a different centre. It is also important to note that the clinical aims of the model, for example, for patient advice or to compare institutions should be considered when assessing any model.

Nilsson *et al.*³ compare 19 risk-score algorithms regarding their validity to predict both 30-day and 1-year mortalities after cardiac surgical procedures. Their study included 6222 cardiac operations performed between 1996 and 2001 at a single Swedish institution, with a prospective data collection. They used receiver-operating characteristic (ROC) curves to describe the performance and accuracy of risk-score algorithms. Actual mortality was 2.9% at 30 days and 6.1% at 1 year. For overall open-heart surgery, Nilsson *et al.*³ report that discriminatory power for 30-day and 1-year mortalities was highest for logistic (0.84 and 0.77) and additive (0.84 and 0.77) EuroSCORE algorithms. The next highest were Cleveland Clinic (0.82 and 0.76) and Magovern (0.82 and 0.76) scoring systems. Furthermore, in coronary artery bypass grafting (CABG)-only surgery (4351 procedures), EuroSCORE had the highest discriminatory power both for 30-day and 1-year mortalities, followed by New York State (NYS) and Cleveland Clinic risk scores. Nilsson *et al.*³ conclude that although EuroSCORE, Cleveland Clinic, Magovern, and NYS risk scoring models were originally designed to predict early mortality after cardiac surgery, they can also accurately predict 1-year mortality.

Prediction of operative mortality

An area under the curve >70% is usually considered to be associated with a good predictive value.⁵ The finding that EuroSCORE, probably the most commonly used scoring algorithm in Europe, can accurately predict

¹ doi:10.1093/eurheartj/ehi720

operative mortality after open-heart surgery and after CABG-only procedures is not surprising and confirms the results of previous studies.^{1,6,7} Furthermore, despite substantial geographic differences between Europe and North America, EuroSCORE performs very well in the Society of Thoracic Surgeons database. Therefore, it can be considered an adequate risk stratification system on both sides of the Atlantic.⁷

Predictive values for older scoring algorithms are usually poorer when compared with more recent ones. For example, as in Nilsson's³ report, overall mortality is overestimated by the initial Parsonnet score in most studies.^{6,8} The database for the Parsonnet score¹ is now almost 20 years old, and its predictive value has probably be lessened by medical and surgical therapy advances achieved in this period. As this process applies to any scoring systems over time, revalidation of score items at regular intervals is most likely necessary.

Short-term mortality is probably not by itself an adequate indicator of quality of care or resource use. On the opposite, morbidity is a major determinant of hospital cost and quality of life after surgery. Being more frequent than mortality, it could carry more information and be measured in terms of postoperative complications and length of hospital stay. However, for most scoring models, predictive values for morbidity are considerably lower than predictive values for mortality.⁸ Therefore, development of specific morbidity risk scores could improve outcome and hospital cost prediction. Furthermore, because of the heterogeneity of morbidity events, future scoring systems should probably generate separate predictions for mortality and major morbidity events.

Recent studies demonstrated that EuroSCORE could be correlated with costs of cardiac surgery.^{9,10} Pre-operative risk algorithms able to accurately predict cost of care and hospital resource needs are desirable. Such algorithms could tentatively be used to plan the optimal schedule for cardiac surgery, moderate the postoperative workload in the intensive care unit, and rationally allocate hospital resources.¹⁰

Prediction of 1-year mortality

Long-term mortality, probably the most useful outcome, is rarely assessed, essentially because of the difficulty in following patients over a long period of time. It certainly should be a priority for future research in risk modelling. In this regard, the findings by Nilsson *et al.*³ that some risk stratification models, such as EuroSCORE, can predict 1-year mortality is potentially important, but requires further validation.

As Nilsson *et al.*³ correctly pointed out, smaller ROC area is expected for 1-year mortality prediction when compared with short-term mortality. This is because the proportion of cardiovascular deaths, among all causes of mortality, is usually lower at 1 year than at 30 days after surgery.

Limitations

The predictive accuracy of all risk-score algorithms is influenced by numerous factors, including variable definitions, management of incomplete data field, geographic differences in patient risk factors, and surgical procedure selection criteria.

Furthermore, it should be emphasized that accuracy and discriminant power can be fairly independent, as a model that soundly over- or under-estimates the probability of death can be efficient in discriminating patients who will survive from those who will die. According to the Bayes theorem, positive predictive value is very low when the disease prevalence is low, even if the discriminatory test has a high sensitivity and specificity. This partially explains why predictive power of most models is poorer for (very) high-risk patients.

Conclusions

There is a significant variation in patients risk profile and surgical strategy both within Europe and across the Atlantic. Therefore, it is not appropriate to assess the quality of care by measuring crude procedural mortality alone. It emphasizes that comparisons of operative mortality rates among centres are meaningless without risk adjustments derived from casemix. In the United States, in the interest of consumer education, the publication of mortality data in newspapers and other media sources under the guise of allowing the consumers to make a better choice has resulted in denial of cardiac surgery to high-risk patients. If the medical community be unprepared, a similar situation could happen in Europe in the near future.

In this regard, pre-operative risk stratification models are useful tools to compare quality in different centres and to assess costs related to patients severity. Data collection and risk stratification are of paramount importance for proper quality assessment and outcome improvement in cardiac surgery. Risk stratified data are essential for

quality analysis, meaningful comparison of outcomes, and improvements of outcomes. It should be an integral part of the cardiac surgical practice, being as essential to the surgeon as the knowledge of anatomy and techniques. It belongs to risk assessment, decision-making, and informed consent.

However, cardiologists and cardiac surgeons should bear in mind that when using predictive models at bedside to provide the patient with an estimate of surgical risk, they assign a reliable probability of death of a population and not for the actual patient. It should also be underlined that risk stratification models score the risks of care, but not the quality of care.

Conflict of interest: none declared.

References

1. 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; 79:13-112.
2. Higgins TL. Quantifying risk and assessing outcome in cardiac surgery. *J Cardiothorac Vase Anesth* 1998; 12:330-340.
3. Nilsson J, Algotsson L, Höglund P, Lührs C, Brandt J. Comparison of 19 pre-operative risk stratification models in open-heart surgery. *Eur Heart J* 2006;27:867-874. First published on January 18, 2006, doi:10.1093/eurheartj/ehi720.
4. Eagle KA, Guyton RA, Davidoff R, Edwards FH, Ewy GA, Gardner TJ, Hart JC, Herrmann HC, Hillis LD, Hutter AM Jr, Lytle BW, Marlow RA, Nugent WC, Orszulak TA. 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-e437.
5. Swets JA. Measuring the accuracy of diagnostic systems. *Science* 1988; 240:1285-1293.
6. Asimakopoulos G, Al Ruzzeh S, Ambler G, Omar RZ, Punjabi P, Amrani M, Taylor KM. 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;23:935-941.
7. Nashef SA, Roques F, Michel P, Cortina J, Faichney A, Gams E, Harjula A, Jones MT. Coronary surgery in Europe: comparison of the national subsets of the European system for cardiac operative risk evaluation database. *Eur J Cardiothorac Surg* 2000; 17:396-399.
8. Geissler HJ, Holzl P, Marohl S, Kuhn-Regnier F, Mehlhorn U, Sudkamp M, de Vivie ER. Risk stratification in heart surgery: comparison of six score systems. *Eur J Cardiothorac Surg* 2000; 17:400-406.
9. Pinna PP, Bobbio M, Colangelo S, Veglia F, Marras R, Diena M. Can EuroSCORE predict direct costs of cardiac surgery? *Eur J Cardiothorac Surg* 2003; 23:595-598.
10. Nilsson J, Algotsson L, Höglund P, Lührs C, Brandt J. EuroSCORE predicts intensive care unit stay and costs of open heart surgery. *Ann Thorac Surg* 2004; 78:1528-1534.