The EuroSCORE risk stratification system in the current era: how accurate is it and what should be done if it is inaccurate?

Keywords: Logistic EuroSCORE; Risk stratification model; Cardiac; Coronary artery bypass surgery; Recalibration

The study by Zheng et al. on 'The Chinese Coronary Artery Bypass Grafting Registry Study: how well does the EuroSCORE predict operative risk for Chinese population?' has found an overestimation of the predicted in-hospital mortality using the logistic EuroSCORE risk stratification model [1]. In their study, Zheng et al. have evaluated 9248 cardiac surgical patients who underwent surgery between January 2004 and December 2005. The predicted mortality was 5.51% while the observed mortality was 3.27%. In the subgroup of 8120 patients who had isolated coronary artery bypass grafting (CABG) surgery, the predicted mortality by logistic EuroSCORE model was 4.21% while the observed mortality was 2.22%. Discrimination was tested by determining the area under the receiver operating characteristic (ROC) and this was 0.72 for the entire cohort and 0.71 for the isolated CABG subset. In view of these findings, the authors concluded that the logistic EuroSCORE model over-predicts in-hospital mortality and therefore does not accurately predict outcomes in this group of Chinese CABG patients [1].

This article is timely and highlights a number of important points:

1. The logistic EuroSCORE risk stratification system was developed and validated within the European population [2,3]. There should be caution in the utilisation of any particular risk stratification system outside the countries of origins, and it is important to carefully evaluate the validity of such system amongst foreign population [4–6].

2. The additive EuroSCORE risk model was developed utilising data from 14781 patients from 128 surgical centres in 8 European states who underwent surgery between September and November 1995, and was published in 1999 [7,8]. Subsequently the logistic EuroSCORE model was developed to provide a better risk predictor, especially in high-risk patients and was published in 2003 [2,3]. The model is now 6 years old and was developed based on 14,799 patients from the EuroSCORE database who had surgery in 1995, and hence the model may be outdated [2]. Since 1995, there have been improvements in perioperative care and surgical techniques, all of which have led to improved surgical results and a reduction in hospital mortality. The information that logistic EuroSCORE over-predicts mortality is not new and several recent studies have found that [5,6,9]. 'The Italian CABG Outcome Project' analysed 30,610 isolated CABG and found that the logistic EuroSCORE predicted mortality was 6.27% while the observed mortality was 2.54% [6]. When patients were divided into six risk classes, the logistic EuroSCORE continue to overestimate mortality by a factor of approximately 2.5 in all classes [6]. The 'North West Quality Improvement Programme in Cardiac Interventions' in the United Kingdom (UK) evaluated 9995 patients who underwent surgery and found that although the discrimination of the logistic EuroSCORE was good with ROC curve area of 0.79 for all cardiac surgery, it however over-predicted in-hospital mortality [9]. The predicted mortality was 5.7% while the observed mortality was 3.3%. Analyses of subgroups based on types of surgery found that logistic EuroSCORE over-predicted observed mortality for all subgroups and by differing degrees [9]. Similar observations were seen by 'The Society for Cardiothoracic Surgery in Great Britain and Ireland' on analyses of the UK national database.

3. The information that the logistic EuroSCORE model presently over-predicts in-hospital mortality highlights the importance of vigilant ongoing evaluation of any risk stratification model, and systems should be set in place to detect risk models that may have become less accurate over time.

4. Optimal medical care is a finely tuned balance between risk and benefit. It stands in close relation to the variability of the patient and to the socio-economic environment wherein this care is delivered. The variability of the patient consists of a series of tacit and codified elements. Body mass index can be codified easily but physical and mental reserves are very tacit elements. An appropriate correction for variability demands codification of all variability. Socio-economic and scientific limitations however at times restrict this documentation, and therefore the theoretical possibility of this correction process. The EuroSCORE uses a restricted list of variables. In addition, the continuous variables are mostly used in nominal expression and are not transformed, thereby reducing the appropriate correction. The major improvement between the additive and logistic score is the refinement in the coefficients used. Exceptional variability is also another limitation. The correction for any certain variability will be most valid around the mean of that variability, dramatically decreasing, once one or two standard deviations are reached. Body mass index, for example, is not included in the EuroSCORE dataset, because it did not influence the risk within the observed
interval and for the variability present in the study population. Although cardiac surgery is only exceptionally done when this value reaches 45 or inversely 15, the risks in such patients are however high. Additionally, the scientific observation visible in the 95% confidence limits (CL) between the predicted and the observed hospital mortalities, as in this publication by Zheng, limits inference building [1]. The 95% CL of the predicted mortality ranges from 4.2% to 7.3% while that of the observed mortality ranges from 2.3% to 4.6%. An extremely large cohort is indeed needed for analysis to achieve a better CL when hospital risk in the current era of cardiac surgery is as low as this. The limitations of any risk stratification systems must be understood and carefully taken into account.

5. The time-span of outcome evaluation is conceptually the complete life expectancy of the patient. Therefore ranking care demands very long follow-up intervals and complete registration of all variability. Early risk is the interaction between the morbidity and the comorbidity of the patient on the one side and the therapeutic process on the other side. Conceptually, this will vary for each therapeutic process. Therefore grouping procedures will widen the uncertainty of any calculation. Also the wrong or incomplete therapeutic process could be associated with differences in short- and long-term outcomes. For example, the outcome differences between total venous CABG versus arterial CABG. The time-spans in the evaluation of outcomes are often under measured. The development of postoperative stroke is generally within the first 8 to 10 days, however the time-span for early mortality may extend at least 3 months for CABG and up to 6 months for valve patients. Early mortality can increase significantly in the first several months versus within the first month, especially in patients with significant comorbidity and postoperative complications. The observations of the whole time-span of these various interactions are therefore important. The EuroSCORE observes only a short and possibly biased interval. This bias will vary according to the socio-economic environment, and also to the institutional procedures. Therefore excluding a large portion of the risk intervals will further increase this uncertainty. The reporting of outcomes needs to expand beyond hospital or 30-day mortality, to include longer time-spans so as to provide a more balanced measurement and reporting.

6. There are variations in case mix between both hospitals and surgeons, and risk stratification systems are essential to adjust for these differences so as to allow for meaningful comparisons. These systems form an integral audit tool for the monitoring of performance of surgeons and hospitals. Reliance on a risk stratification system that over-predicts mortality may have serious consequences. It leads to a false sense of reassurance, underperformance may go undetected, and patient welfare may be compromised. In order to address the issue of over-prediction, some authors have proposed recalibration of the logistic EuroSCORE to provide a more accurate prediction [6, 9]. ‘The Italian CABG Outcome Project’ has found that the logistic EuroSCORE overestimation remains constant through the six risk classes of isolated CABG analysed, with an observed over-predicted mortality ratio of 0.4 [6]. They have concluded that when properly recalibrated, the logistic EuroSCORE model can be exported to the Italian population and used to rank hospital performance and evaluate preoperative risk of patients undergoing open-heart surgery. ‘The North West Quality Improvement Programme in Cardiac Interventions’ has found different factors of over-predictions with different types of surgery [9]. They have proposed different recalibrations for different operative groups [9]. ‘The Society for Cardiothoracic Surgery in Great Britain and Ireland’ have responded to the over-prediction of logistic EuroSCORE by undertaking a complex recalibration whereby they have looked at the comparisons between the observed mortality and that predicted in each operative group to derive a series of recalibration coefficients. These were then applied to the analyses of national data for hospitals and surgeons, which are available to the public (http://heartsurgery.healthcarecommission.org.uk/). This was first done for publication in 2007 based on data from 2002 to 2005. At Papworth Hospital, a recalibration ratio of 0.5 of the logistic EuroSCORE is presently being utilised for internal monitoring of performance. The usage of recalibration coefficients presently allows for temporary adjustments of the over-predictions obtained from logistic EuroSCORE. The time has now arrived for a careful and vigorous revision of the logistic EuroSCORE or the development of a new risk stratification system that utilises data from a contemporary cohort of cardiac surgical patients that will accurately predict mortality in the current era. Considerations should also be given to the development of procedure specific risk models. The reporting of outcomes also needs to expand to include other outcome measures (examples: re-exploration, stroke, renal failure, infection etc.), and not just mortality alone so as to provide a balanced measurement and reporting of outcomes.

In conclusions, all available systems correcting for variability have a number of limitations. Some can be improved, while others are not amendable for scientific or socio-economic reasons. Risk stratification scoring systems are interesting and important rulers to be used in daily practice, but their limitations should be known and included in inference building. The article by Zheng et al. is therefore timely and highlights a number of very important issues that require careful consideration and attention.

References


Cliff K. Choonga,b*, Paul Sergeantc, Samer A.M. Nashefa
Julian A. Smithd, Ben Bridgewatere,f

aDepartment of Cardiothoracic Surgery, Papworth Hospital NHS Foundation Trust, Cambridge CB23 3RE, UK
bDepartment of Surgery, The University of Cambridge, Cambridge, UK
cCardiac Surgery Department, Leuven University Hospital Gasthuisberg, Leuven B-3000, Belgium
dDepartment of Cardiothoracic Surgery, Monash Medical Centre, Southern Health, Monash University, Melbourne, Victoria, Australia
eUniversity Hospital of South Manchester, Southmoor Road, Manchester M23 9LT, UK
fNational Institute for Clinical Outcomes Research, University College of London, UK

*Corresponding author. Address: Papworth Hospital, The University of Cambridge, Cambridge CB23 3RE, UK. Tel.: +44 1480 364573 E-mail address: cliffchoong@hotmail.com (C.K Choong)