Review Article

Predictive risk models for proximal aortic surgery

Daniel Hernandez-Vaquero, Rocío Díaz, Isaac Pascual, Rubén Álvarez, Alberto Alperi, Jose Rozado, Carlos Morales, Jacobo Silva, César Morís

Heart Area, Central University Hospital of Asturias, Oviedo, Spain

Contributions: (I) Conception and design: D Hernandez-Vaquero, R Álvarez, A Alperi, J Rozado; (II) Administrative support: R Díaz, I Pascual; (III) Provision of study materials or patients: None; (IV) Collection and assembly of data: R Díaz, I Pascual; (V) Data analysis and interpretation: C Morales, J Silva, C Morís; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Daniel Hernandez-Vaquero, MD, PhD. Heart Area, Central University Hospital of Asturias, Oviedo, Spain. Email: dhvaquero@gmail.com.

Abstract: Predictive risk models help improve decision making, information to our patients and quality control comparing results between surgeons and between institutions. The use of these models promotes competitiveness and led to increasingly better results. All these virtues are of utmost importance when the surgical operation entails high-risk. Although proximal aortic surgery is less frequent than other cardiac surgery operations, this procedure itself is more challenging and technically demanding than other common cardiac surgery techniques. The aim of this study is to review the current status of predictive risk models for patients who undergo proximal aortic surgery, which means aortic root replacement, supracoronary ascending aortic replacement or aortic arch surgery.

Keywords: Decision support models; aortic aneurysm; thoracic; thoracic surgery

Submitted Jan 07, 2017. Accepted for publication Jan 27, 2017.
doi: 10.21037/jtd.2017.03.91
View this article at: http://dx.doi.org/10.21037/jtd.2017.03.91

Ways of measure the performance of predictive risk models

Discrimination is the ability to distinguish or separate those patients who will develop the event of interest from those who will not. Therefore, it is crucial when the aim is to classify or separate patients at low, medium or high-surgical risk. Calculation of the area under the receiving operating curve (AROC) is the most preferable discrimination measure. This value, also called "C-statistic", is the probability that the predictive model gives a higher value for a random patient who will develop the event of interest (postoperative mortality) than for a random patient who will not. Hence, C-statistic of 0.5 indicates no predictive ability while a value of 1 means perfect discrimination power. So, usually these values are accepted: AROC =0.5 no discrimination accuracy; AROC =0.5–0.6 very poor discrimination accuracy; AROC =0.6–0.7 poor discrimination; AROC =0.7–0.8 fair discrimination accuracy; AROC =0.8–0.9 good discrimination accuracy; AROC >0.9 excellent discrimination accuracy (1-3). Two ROC curves can be compared in order to know which model has better discrimination accuracy. Calculations to do that were described by Hanley and McNeil (4).

Discrimination is one of the most important measures but this cannot be the only consideration. Taking into account two examples already given by Woodward (1), we can consider two risk scores, one of them gives always a mortality risk value 0.3 higher than the other. AROC would be exactly the same but logically there are great differences between both system predictions. Similarly, a test that assigns a mortality risk of 0.4 to all patients who will die and 0.2 to all patients who will survive would have the same discrimination accuracy than other model giving a mortality risk of 0.99 for all patients who will die and 0.01 for all patients who will survive. Logically, the second model is much better but discrimination accuracy would be the same. Other example was given by Cook (2), a model that assigns all cases a value of 0.52 and all noncases a value of 0.51 would have perfect discrimination, although these probabilities are not helpful. Other problems regarding the use of the C-Statistic have been reported by the same author (2).

Calibration is the agreement between the predicted and observed risks. For instance, if a model predicts a 20% mortality rate, the observed mortality rate should be 20 out of 100 patients. Grouped calibration is the best way to measure the calibration of a given model. Observed and predicted mean risks are evaluated in several mutually exclusive subgroups (usually 10), which are created according to predicted risks (1). The Hosmer-Lemeshow test is the most frequent way to measure grouped calibration. Limitations as its dependence on the sample size have been attributed. In general terms, if p value is under 0.05 the test shows bad calibration but if p value is over 0.05 it only means that the model does not find bad calibration, which does not mean that it is good, but only that the model is not able to demonstrate bad calibration. So, calibration accuracy will be found statistically not significant if the sample size is small enough Similarly, the power increases with sample size; this can be undesirable for goodness of fit tests because in very large data sets, small departures from the proposed model will be considered significant. Some ways of reducing this problem, as modifying the number of subgroups, have been published (5). Moreover, this test is not able to say how good or how bad the calibration is. Other concerns on the use of this test have been reported elsewhere (6).

The risk adjusted mortality ratio (RAMR) is other measure for calibration. This calculation, which is the ratio between the expected and observed mortality (E/O), is useful to know if the system overestimates of underestimates the real surgical risk. So, a ratio over 1 means that the system overestimates the real surgical risk while under 1 the model would underestimates it. Assuming the Poisson distribution, the confidence intervals of RAMR may be calculated using this formula: [IC 95%= E/O*exp(+/–1.9596*O^–0.5)] (3,7). Calibration is not bad if the 1 value is inside this confidence interval and vice versa.

Other ways to measure calibration have been described elsewhere (1). There is no perfect test for measuring calibration but there is agreement that descriptive plots generally offer the best information.

Using calibration and discrimination we can evaluate how well or bad a model is able to predict. However, it would be useful a test which evaluates the overall model performance. Ranging from 0 for a perfect system to 0.25 for a model with no predictive power, the Brier score is the most useful way to measure overall model predictive power. The formula is simply the average of the squared differences between the observed and predicted risks (1).

How do EuroSCORE I and EuroSCORE II predict mortality risk after proximal aortic surgery?

In 1995, the European system for cardiac operative risk evaluation (EuroSCORE) was published as a risk model derived from data collected from 14,799 consecutive patients in 100 European centers (8). As the years went by and the external validation was assessed, EuroSCORE I showed to overestimate the mortality risk of the patients undergoing cardiac surgery (9).

Some reasons have been argued. First and very important, the system was developed using data from a highly heterogeneous population. Thus, because most of the group had coronary artery disease, EuroSCORE I was more predictive for patients undergoing coronary artery surgery. Second, the model only took into account the most frequent risk factors but not for rare ones as cirrhosis or porcelain aorta. Finally, surgical techniques and postoperative care improved over time making EuroSCORE I a redundant model (9).

With the aim of overcoming these limitations, EuroSCORE II was born in 2012 (10). More than 22,000 patients were analyzed collecting data from 154 hospital centers, most of them, from Spain. When external validation was assessed, some authors (11-13) found that this system did not accurately predict in some groups of patients showing that the accuracy of any predictive model depends largely on the homogeneity between the population used to create the system and the study sample.

Although surgical activity in proximal aortic operations is relatively low compared with other cardiac surgery operations (14), the procedure itself carries a greater risk of mortality so predictive risk models for aortic surgery are of utmost importance (15). These systems help improve decision making, informed consent to our patients and quality control comparing results between surgeons and between institutions. The use of these models promotes competitiveness and led to increasingly better results (16).

Since STS on-line calculator does not address aortic surgery, the performance of EuroSCORE I and EuroSCORE II is of utmost importance in the cardiothoracic surgical community.

Nishida *et al.* are probably the most experienced authors on this issue. In 2006, they reported the superiority of the logistic EuroSCORE over the additive EuroSCORE in predicting mortality after aortic surgery in 327 Japanese patients during a 30-year period at their institution (17). In 2013 they compared EuroSCORE II with its previous

versions in a cohort of 461 patients who underwent aortic surgery during a period of 20 years (18). Although they did not compare both ROC curves in a statistical way (4), they found a C-statistic higher for EuroSCORE II (AROC =0.77) suggesting that this system was able to predict better than the logistic EuroSCORE I (AROC =0.72). Unfortunately, specific calibration measures as aforementioned were not calculated but they suggested that logistic EuroSCORE failed mainly in high-risk patients.

Using the database of the Dutch Committee of Heart Interventions, Huijskes *et al.* (19) analyzed 1,290 patients who underwent thoracic aorta surgery in Netherlands to find that discriminatory power of logistic EuroSCORE was poor (AROC =0.64). No calibration measure was reported. However, when external validation fails, often the problem lies with calibration rather than discrimination (1).

Barmettler *et al.* (20) studied external validation of EuroSCORE in 367 patients who underwent surgery on the thoracic aorta, including type A dissections in their institution. AROC for logistic EuroSCORE was 0.72. This means that 28% of the time (1–0.72), this system gives a lower expected risk for a patient who will die than for a patient who will not. Moreover, no specific calibration measure was reported.

As aforementioned, discrimination cannot be the only measure to study model performance. These studies fail to determine the calibration of EuroSCORE I in this population. However, since discrimination power was not very good, they seem to suggest other systems or models to be evaluated or created. EuroSCORE II must be further investigated calculating appropriate calibration measurements.

New predictive risk models for ascending aortic surgery

In 2012, data of 45,894 patients who underwent proximal aortic replacement were published (15). While elective operative mortality rate was 3.4%, urgent and emergent procedures had worse outcomes with a mortality risk of 8.3% which enhances the need of a reliable risk predictive model.

Risk factors for mortality were analyzed and a new predictive risk model was created. The most important predictor of mortality was emergent surgery with an OR =5.9. Other risk factors included urgent surgery, concomitant coronary artery by-pass grafting (CABG), concomitant mitral valve procedure and surgery with arch involvement. Surprisingly, aortic root replacement, which is technically more challenging, was not a prognostic factor. The authors of this predictive model reported a good discrimination power (C-statistic of 0.82) (15). This model is probably the most powerful predictive system ever created for proximal aortic surgery. Unfortunately, specific calibration measurement was not reported. External validation of the model is required before it can be used worldwide.

Using for first time a large contemporaneous European cardiac surgery database, Bashir *et al.* (21) developed and validated two risk prediction models for postoperative mortality after surgical procedure on the proximal aorta (i.e., root, ascending, or arch segments). One of them was created to be used in non-elective patients and the other in elective procedures. To achieve this formidable challenge they used data for 8,641 consecutive UK patients undergoing proximal aortic operation from the National Institute for Cardiovascular Outcomes Research database from April 2007 to March 2013.

In the elective group, previous cardiac operation was the most important predictor of postoperative mortality. Other risk factors included age, left ventricular dysfunction, surgery on the aortic arch, triple vessel disease and concomitant coronary surgery, neurologic disease, aortic disease other than aneurysm, pulmonary disease, preoperative nonsinus rhythm, female sex and NYHA functional class >II/IV (21).

For non-elective patients, salvage priority was the most important predictor with an OR =9.13; 95% CI (5.93–14.05). Previous cardiac operation was also a strong predictor, followed by emergency priority, concomitant CABG, age, preoperative nonsinus rhythm, cardiogenic shock, creatinine >200 mmol/L, preoperative ventilation and peripheral vascular disease (21).

Calibration based on Hosmer-Lemeshow test and plots showed good overall calibration in both models but the elective system overestimated the real surgical risk for patients with an expected mortality over 40%. Discrimination was quite good with AROC =0.80 for the elective group and 0.76 for the non-elective model. These risk factors in the non-elective group as well as the adequate internal validation are consistent with a study published 14 years ago for patients with type A aortic dissection (22). Studies on external validation of these models (15,21,22) are needed and if they perform well, an on-line calculator could be a final solution.

Conclusions

Although surgical activity in proximal aortic operations is relatively low compared with other cardiac surgery

operations, the procedure itself carries a greater risk of mortality so predictive risk models are of utmost importance (14,15). These systems help improve decision making, informed consent to our patients and quality control comparing results between surgeons and between institutions.

Since on-line calculators for proximal aortic surgery do not exist, generic models as logistic EuroSCORE and EuroSCORE II are of utmost importance for cardiac surgeons. Logistic EuroSCORE seems not to be able to adequately predict the real mortality risk of these patients. There is an urgent need to study EuroSCORE II model performance for proximal aortic surgery using proper tools for discrimination and calibration measures. However, since the accuracy of any predictive model depends largely on the homogeneity between the population used to create the system and the study sample, it is likely that EuroSCORE II does not predict excellently in this population.

It is likely that great variability exists in the results of proximal aortic surgery between surgeons and centers (14). For this reason, it might be difficult to find a good model, able to predict the mortality risk after these challenging and technically demanding procedures. Some researchers have created and published rigorous predictive models for proximal aortic surgery but studies on external validation are needed. In case of getting good external validation, an on-line calculator for these operations would be highly appreciated for surgeons, clinicians and patients worldwide.

Non-elective procedure is probably the most important predictor of postoperative mortality after proximal aortic surgery. Concomitant CABG, aortic arch surgery, age and neurologic dysfunction play also a key role. However, involvement of the aortic root seems not to be an important variable despite requiring a more difficult intervention than supracoronary ascending aortic replacement.

Acknowledgements

None.

Footnote

Conflicts of Interest: The authors have no conflicts of interest to declare.

References

1. Woodward M. Risk scores and clinical decision rules. In: Woodward M. editor. Epidemiology. Study Design and Data analysis. Third Edition. Florida: Chapman & Hall/CRC biostatistics series, 2014:605-78.
2. Cook NR. Use and misuse of the receiver operating characteristic curve in risk prediction. Circulation 2007;115:928-35.
3. Kengne AP, Beulens JW, Peelen LM, et al. Non-invasive risk scores for prediction of type 2 diabetes (EPIC-InterAct): a validation of existing models. Lancet Diabetes Endocrinol 2014;2:19-29.
4. Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. Radiology 1983;148:839-43.
5. Paul P, Pennell ML, Lemeshow S. Standardizing the power of the Hosmer-Lemeshow goodness of fit test in large data sets. Stat Med 2013;32:67-80.
6. Bertolini G, D'Amico R, Nardi D, et al. One model, several results: the paradox of the Hosmer-Lemeshow goodness-of-fit test for the logistic regression model. J Epidemiol Biostat 2000;5:251-3.
7. Boggs DA, Rosenberg L, Pencina MJ, et al. Validation of a breast cancer risk prediction model developed for Black women. J Natl Cancer Inst 2013;105:361-7.
8. Nashef SA, Roques F, Michel P, et al. European system for cardiac operative risk evaluation (EuroSCORE). Eur J Cardiothorac Surg 1999;16:9-13.
9. Hernández-Vaquero D, Díaz R, Morís C. Predictive risk models for transcatheter procedures: how should they be created? J Thorac Cardiovasc Surg 2014;148:1759.
10. Nashef SA, Roques F, Sharples LD, et al. EuroSCORE II. Eur J Cardiothorac Surg 2012;41:734-44; discussion 744-5.
11. Howell NJ, Head SJ, Freemantle N, et al. The new EuroSCORE II does not improve prediction of mortality in high-risk patients undergoing cardiac surgery: a collaborative analysis of two European centres. Eur J Cardiothorac Surg 2013;44:1006-11; discussion 1011.
12. Osnabrugge RL, Speir AM, Head SJ, et al. Performance of EuroSCORE II in a large US database: implications for transcatheter aortic valve implantation. Eur J Cardiothorac Surg 2014;46:400-8; discussion 408.
13. Hernández-Vaquero D, Díaz R, Meana B, et al. External validation of the EuroSCORE II risk stratification model in the USA. Eur J Cardiothorac Surg 2015;48:177.
14. Bustamante-Munguira J, Centella T, Polo L, et al. Cirugía cardiovascular en España en el año 2014. Registro de intervenciones de la Sociedad Española de Cirugía Torácica-Cardiovascular. Cir Cardiov 2015;22:297-313.
15. Williams JB, Peterson ED, Zhao Y, et al. Contemporary results for proximal aortic replacement in North America.

J Am Coll Cardiol 2012;60:1156-62.
16. Nashef SA. The Naked Surgeon: the power and peril of transparency in medicine. London: Scribe Publications, 2015.
17. Nishida T, Masuda M, Tomita Y, et al. The logistic EuroSCORE predicts the hospital mortality of the thoracic aortic surgery in consecutive 327 Japanese patients better than the additive EuroSCORE. Eur J Cardiothorac Surg 2006;30:578-82; discussion 582-3.
18. Nishida T, Sonoda H, Oishi Y, et al. The novel EuroSCORE II algorithm predicts the hospital mortality of thoracic aortic surgery in 461 consecutive Japanese patients better than both the original additive and logistic EuroSCORE algorithms. Interact Cardiovasc Thorac Surg 2014;18:446-50.
19. Huijskes RV, Wesselink RM, Noyez L, et al. Predictive models for thoracic aorta surgery. Is the Euroscore the optimal risk model in the Netherlands? Interact Cardiovasc Thorac Surg 2005;4:538-42.
20. Barmettler H, Immer FF, Berdat PA, et al. Risk-stratification in thoracic aortic surgery: should the EuroSCORE be modified? Eur J Cardiothorac Surg 2004;25:691-4.
21. Bashir M, Shaw MA, Grayson AD, et al. Development and Validation of Elective and Nonelective Risk Prediction Models for In-Hospital Mortality in Proximal Aortic Surgery Using the National Institute for Cardiovascular Outcomes Research (NICOR) Database. Ann Thorac Surg 2016;101:1670-6.
22. Mehta RH, Suzuki T, Hagan PG, et al. Predicting death in patients with acute type a aortic dissection. Circulation 2002;105:200-6.

Cite this article as: Hernandez-Vaquero D, Díaz R, Pascual I, Álvarez R, Alperi A, Rozado J, Morales C, Silva J, Morís C. Predictive risk models for proximal aortic surgery. J Thorac Dis 2017;9(Suppl 6):S521-S525. doi: 10.21037/jtd.2017.03.91

