

The business of risk



Ryan J. Lennon, MS; David R. Holmes Jr*, MD

Mayo Clinic, Rochester, MN, USA

Predictions about the future feature prominently in all aspects of human endeavour from weather to economics to politics and, increasingly, to medicine. In cardiovascular disease where scientists often deal with strategies of care, the design of clinical trials is greatly influenced by consideration of risk in the patients to be treated; in this case, issues of patient population, baseline demographics and predicted event rates are used to design sample size and power calculations. For patients, risk prediction is less ethereal, instead being very concrete and of the utmost importance because the potential endpoints to be predicted are often hard ones and once they occur they cannot be taken back, such as stroke, myocardial infarction or mortality.

In the field of structural heart disease, transcatheter aortic valve replacement (TAVR) occupies a position of great importance by virtue of treating what had been an unmet clinical need, the number of patients who have actually been treated, the fact that there are alternative treatments available, namely surgical aortic valve replacement (SAVR), the potential for further expansion of the instructions for use (IFU), and the fact that associated patient comorbidities greatly increase risk so that hard points, such as mortality and stroke, are of significant concern and more than rare.

In this issue of EuroIntervention, Arsalan et al¹ provide important information for the field by using their unique data set of 946 consecutive patients undergoing TAVR to validate the recently developed STS/ACC TAVI risk score for in-hospital TAVR mortality and compare its ability to predict 30-day mortality with that of four other established risk models, including EuroSCORE I², EuroSCORE II³, STS-PROM⁴, and the German AV Score⁵.

Article, see page 1520

Although EuroSCORE I and II and STS-PROM are widely used, they were designed and tested in patients undergoing conventional cardiac surgery, whereas the German Registry included both TAVR and SAVR. Accordingly, since many TAVR patients are felt to be at either high or even prohibitive risk, the relevance of these risk scores to TAVR has been uncertain. There are other risk scores which have been developed which have focused on TAVR patients, including the TAVI2-SCORE⁶, the FRANCE-2 score⁷, and the GARY risk score⁸. There have been issues with each of these, including limited sample sizes, and lack of validation on a different data set. This increasing number of risk scores attests to the importance of the development of improved scores to help optimise patient selection, and to providing education for patients and their families about the risk/benefit considerations of TAVR.

*Corresponding author: Mayo Clinic, 200 First Street SW, Rochester, MN 55905, USA.

E-mail: holmes.david@mayo.edu

Comparative analyses of risk scores are complex, as patient populations used for development may vary, techniques of deployment may vary, and technology continues to evolve. The baseline characteristics of patients in this German registry were similar to those seen in other registries⁹. The median age was 82.1 years, approximately 50% were female, 20.4% had COPD and 14.6% had a prior history of stroke. The median predicted baseline risk score varied from as low as 3.7 with the German AV Score, to 5.0 seen with both STS-PROM and the EuroSCORE II, and 21.1% with the EuroSCORE I. The majority of patients were felt to be at high or intermediate risk by the German team and were representative of current clinical practice in Germany. During the initial hospitalisation, 48 patients (4.9%) died, while at 30 days the mortality rate was 6.3% (60/946).

When assessing the numerical qualities of a risk score, there are generally two concepts to keep in mind – “calibration” and “discrimination”. Calibration refers to whether the expected number of events tends to be equal to the actual number of observed events. Discrimination refers to whether patients with a higher predicted risk are actually more likely to suffer events than patients with a lower predicted risk, regardless of whether those predictions are accurate. As an analogy, consider three different meteorologists attempting to predict rain in a city that gets rain every Monday and Friday, but never on any other day. The first meteorologist notices that it rains two days out of every week, and thus predicts a 28% chance of rain every day. This meteorologist is properly calibrated, but has no discriminatory ability. The second meteorologist notices the pattern but is nervous that the viewers might be caught without an umbrella if the pattern changes, and thus predicts 100% chance of rain for every Monday and Friday, and a 60% chance of rain for the other five days of the week. This meteorologist has perfect discrimination, since the rainy days are always accompanied by a higher predicted risk, but is not well calibrated, since their predictions average to five days of rain per week. Only the third meteorologist, who predicts 100% chance of rain for Mondays and Fridays, and 0% chance of rain every other day, has perfect calibration and discrimination.

For a patient and physician who are assessing risk in the hope of making a treatment decision, calibration would be of primary importance. Even if the score cannot give a very individualised estimate, one would hope that the predicted risk is generally unbiased. Table 2 of the paper demonstrates that the estimated risks from the EuroSCORE I (median risk of 21%) are not suitable for the study population which had an observed 30-day mortality of 6.3%. The other scores seem quite reasonable with median risks of 3.7% to 5.0%. We would expect the median to be slightly below the observed mortality rate, since there will be some very large predicted risks which would raise the mean predicted risk above the median. Additionally, the Hosmer-Lemeshow plots in the paper (Figure 3) are also helpful for this assessment. Unfortunately, the authors re-calibrated all the scores before this analysis, so that readers cannot assess the accuracy of the actual published risk estimates, but rather only the revised estimates. The plot for the STS/ACC

score in Figure 3 shows some issues with the score, even when using re-calibrated estimates. Consider the two right-most points on the plot. These indicate that there are two sets of patients with a similar predicted risk (about 12%), yet the actual observed mortality rate was about 10% for one group, and approximately 17% for the other. Patients and physicians might not agree that 17% and 10% are similar groups of risk, and would certainly prefer a risk score that assigns differential levels of risk to those two groups. Similarly, the left-most point indicates a set of patients who had a median predicted risk around 3%, which is higher than two other groups, yet had no events at all. Thus, while the plot indicated a general agreement between observed and predicted 30-day mortality, room for improvement with regard to calibration remains.

The primary measure of discrimination in the paper by Arsalan et al¹ is the C-statistic, which can be interpreted as follows. Suppose two patients are randomly selected from the sample, one of whom died within 30 days and the other survived to 30 days. The C-statistic is the probability that the risk score appropriately assigned a higher risk to the patient who died. A flipped coin will succeed 50% of the time (a C-statistic of 0.50). The STS/ACC score achieved a C-statistic of 68%, which is certainly better than a coin flip, and was substantially better than the EuroSCORE I (0.55) and EuroSCORE II (0.58). It was slightly better than the German AV Score (0.62), though apparently not reaching statistical significance, and was similar to the STS-PROM score (0.68). When considering whether a C-statistic of 0.68 represents acceptable performance, it may be worthwhile to speculate how an experienced cardiovascular interventionalist might perform if asked to identify which of two random patients is truly at higher risk. Some authors have considered that the C-statistic should exceed 0.7 to be satisfactory.

A third important property of any risk score is the concept of parsimony. An extensive list of candidate variables is available for formulation of any specific risk score. Exhaustive lists render scores too cumbersome and would result in limited use in clinical practice. While the STS-PROM and the STS/ACC TAVI risk scores had similar performance, the former requires 28 variables and the latter only 12. This may potentially result in more widespread adoption of the STS/ACC score, with no apparent loss in model performance relative to the other available scores.

The development and successful practice of TAVR remains a significant achievement for cardiovascular medicine, by providing treatment for patients previously deemed to be too high risk for conventional surgery. Now, 15 years after the initial clinical experience, the task of identifying the highest risk of these high-risk patients remains a challenge. Future work should seek to improve on the discriminatory ability of current risk scores, as well as investigate the accuracy of the published risk estimates in various populations. Until then, physicians should continue to assist patients with decision making using the best available data, while remaining aware of the limitations of these statistical models.

Conflict of interest statement

The authors have no conflicts of interest to declare.

References

1. Arsalan M, Weferling M, Hecker F, Filardo G, Kim WK, Pollock BD, Van Linden A, Arsalan-Werner A, Renker M, Doss M, Kalbas S, Hamm CW, Liebetrau C, Mack MJ, Walther T. TAVI risk scoring using established versus new scoring systems: role of the new STS/ACC model. *EuroIntervention*. 2018;13:1520-26.
2. Roques F, Michel P, Goldstone AR, Nashef SA. The logistic EuroSCORE. *Eur Heart J*. 2003;24:881-2.
3. Nashef SA, Roques F, Sharples LD, Nilsson J, Smith C, Goldstone AR, Lockowandt U. EuroSCORE II. *Eur J Cardiothorac Surg*. 2012;41:734-44; discussion 744-5.
4. O'Brien SM, Shahian DM, Filardo G, Ferraris VA, Haan CK, Rich JB, Normand SL, DeLong ER, Shewan CM, Dokholyan RS, Peterson ED, Edwards FH, Anderson RP; Society of Thoracic Surgeons Quality Measurement Task Force. The Society of Thoracic Surgeons 2008 cardiac surgery risk models: part 2--isolated valve surgery. *Ann Thorac Surg*. 2009;88:S23-42.
5. Kötting J, Schiller W, Beckmann A, Schafer E, Döbler K, Hamm C, Veit C, Welz A. German aortic valve score: a new scoring system for prediction of mortality related to aortic valve procedures in adults. *Eur J Cardiothorac Surg*. 2013;43:971-7.
6. Debonnaire P, Fusini L, Wolterbeek R, Kamperidis V, van Rosendaal P, van der Kley F, Katsanos S, Joyce E, Tamborini G, Murtatori M, Gripari P, Bax JJ, Marsan NA, Pepi M, Delgado V. Value of the "TAVI2-SCORE" versus surgical risk scores for prediction of one year mortality in 511 patients who underwent transcatheter aortic valve implantation. *Am J Cardiol*. 2015;115:234-42.
7. Iung B, Laouenan C, Himbert D, Eltchaninoff H, Chevrel K, Donzeau-Gouge P, Fajadet J, Leprince P, Leguerrier A, Lievre M, Prat A, Teiger E, Laskar M, Vahanian A, Gilard M; FRANCE 2 Investigators. Predictive factors of early mortality after transcatheter aortic valve implantation: individual risk assessment using a simple score. *Heart*. 2014;100:1016-23.
8. Leon MB, Smith CR, Mack MJ, Makkar RR, Svensson LG, Kodali SK, Thourani VH, Tuzcu EM, Miller DC, Herrmann HC, Doshi D, Cohen DJ, Pichard AD, Kapadia S, Dewey T, Babaliaros V, Szeto WY, Williams MR, Kereiakes D, Zajarias A, Greason KL, Whisenant BK, Hodson RW, Moses JW, Trento A, Brown DL, Fearon WF, Pibarot P, Hahn RT, Jaber WA, Anderson WN, Alu MC, Webb JG; PARTNER 2 Investigators. Transcatheter or surgical aortic-valve replacement in intermediate-risk patients. *N Engl J Med*. 2016;374:1609-20.
9. Grover FL, Vemulapalli S, Carroll JD, Edwards FH, Mack MJ, Thourani VH, Brindis RG, Shahian DM, Ruiz CE, Jacobs JP, Hanzel G, Bavaria JE, Tuzcu EM, Peterson ED, Fitzgerald S, Kourtis M, Michaels J, Christensen B, Seward WF, Hewitt K, Holmes DR Jr; STS/ACC TVT Registry. 2016 Annual Report of The Society of Thoracic Surgeons/American College of Cardiology Transcatheter Valve Therapy Registry. *J Am Coll Cardiol*. 2017;69:1215-30.