The risk and reward of surgical aortic valve replacement

Joseph E. Bavaria, MD

Risk assessment has come to be regarded as an essential element in the practice of cardiothoracic surgery. This has become even more critical in the current era when patients have numerous surgical- and transcatheter-based options for therapy. Therefore, I read with great interest the article by Johnston and colleagues, which presents their excellent single-institution operative results in low-risk patients undergoing surgical aortic valve replacement (SAVR) for the 2005-2016 period. The authors compared their observed surgical outcomes to the raw risk estimates (observed to expected [O:E]) derived from the 2008 Society of Thoracic Surgeons Adult Cardiac Surgery Database (STS-ACSD) risk models. The authors concluded: “STS risk models overestimate contemporary SAVR risk at a high-volume center, supporting efforts to create a more agile quality assessment program. SAVR in low-risk patients provides durable survival benefit, supporting early surgery and providing a benchmark for transcatheter AVR.” Although there are several important issues to discuss regarding their reference to 2008 STS risk models, I fully agree with the authors’ statement that the durable survival benefit of isolated SAVR in low-risk patients provides a true benchmark for transcatheter aortic valve replacement (TAVR). The cardiac surgery community should strive to highlight this point more effectively. To further this, I will discuss 5 factors that may provide clarity, including the evolution of the STS Risk Models, the well-documented volume-outcome relationship in cardiac surgery, evolving role of TAVR in low-risk patients, weaponization of quality improvement data, and access to care in cardiac surgery.

EVOLOVU OF THE STS RISK MODELS

The STS-ACSD pioneered the field of surgical risk models and remains the gold standard in cardiac surgery. In 2008, the STS released 27 Risk Models that have developed into the models commonly used today, including AVR with a standardized set of outcomes for all procedures (i.e., mortality, stroke, reoperation, renal failure, deep sternal wound infection, prolonged ventilation, composite major morbidity, prolonged length of stay, and short length of stay). Periodically these models are completely revised, as in 2018, to include new risk factors and to reflect changes in cardiothoracic practice and outcomes. The reference to the 2008 Risk Models in the article by Johnston and colleagues is thus inaccurate. Similarly, to facilitate clinical decision support and patient counseling, the STS recently published an agile and user-friendly online risk calculator that relieves users from the burden of having to apply recalibration factors to adjust raw risk estimates based on model coefficients estimated during the original model development. This calculator will be updated every 3 months to use the latest applicable model coefficients. This cadence of updating the STS online risk calculator accounts for shifting risk across time. Adult cardiac programs and surgeons should feel reassured that all risk estimates used by STS to benchmark their program’s performance and reported to them quarterly (or publicly reported if a program participates) are always based on rederived (i.e., recalibrated) risk models based on the latest 3 years of data or the harvest window.

The STS Risk Models have profound implications, including quality assessment and improvement, program evaluation, patient prognostication, and clinical trial development. The models are based on logistic regression analysis and display excellent calibration with good discriminatory ability that is enhanced in part by the multi-center data collection process. Whereas the very granular details of patient care that may influence outcomes cannot be realistically collected in a national database, there is
rapidly expanding interest in machine learning (ML) in health care, and ML may have a role in predictive analytics in cardiac surgery. Several studies evaluating ML algorithms and other artificial intelligence approaches have demonstrated modest improvements in discriminatory capability for most outcomes compared with existing STS Risk Models. The potential role of ML and artificial intelligence in predictive analytics in cardiac surgery will likely continue to be investigated but do not currently supplant the conventional regression models that have been used and proven to be highly reliable and discriminatory for calculating risk. However, identifying additional data elements that may be predictive of outcomes, in addition to exploring the utility of supplementing ML and logistic regression-based models, may represent mechanisms to improve overall risk prediction and STS has been actively pursuing this.

VOLUME-OUTCOME RELATIONSHIP IN CARDIAC SURGERY

A large body of literature has documented a volume-outcome relationship with numerous operations in cardiac surgery, including a recent study of nearly a quarter-million patients undergoing AVR. This effect is likely multifactorial and related to surgeon experience, system-level care pathways, and infrastructure to prevent failure to rescue. Within the STS risk model, careful and complete descriptions of cardiac surgery outcomes across all program volume groups is critical.

It is important to note that appropriately benchmarked, excellent risk-adjusted outcomes are observed across the entire spectrum of program volumes (small, medium, large, and very large) and with comparable frequency. Similarly, worse-than-expected performance may be seen across the spectrum of program volumes. These observations are demonstrated by an analysis of the 2018-2022 isolated SAVR experience from 850 US programs averaging ≥4 cases per year and a minimum of 20 isolated SAVR cases over this period (Figure 1). It is important to note that volume alone is not a metric by which to assess quality, as highlighted by Thourani and colleagues who observed that prior AVR outcomes are more predictive of future AVR outcomes than a center’s prior AVR volume. These findings highlight the importance of center-level quality improvement efforts to drive longitudinal benefits in surgical outcomes, as suggested by Johnston and colleagues. This concept of benchmarking composite measures in cardiac surgery is not novel and has been a central theme of the STS Quality Measurement Task Force, which has been the driving force behind the STS Risk Models. Therefore, whereas high-volume centers have increased resources and infrastructure to provide robust quality improvement programs, volume alone does not equate to quality and should not be used as a surrogate marker.

EVOLVING ROLE OF TAVR IN LOW-RISK PATIENTS

In the United States, TAVR is approved for treatment of severe aortic stenosis in low-risk patients aged 65 years and older. However, the lack of long-term durability data for TAVR valves have given some surgeons and heart teams pause when considering a transcatheter-first approach in patients with the potential to live 20 or 30 more years. There is emerging evidence that TAVR explant is not a benign procedure and may carry higher risk than redo sternotomy for redo SAVR. These findings highlight the importance of data-driven individualized patient decisions for lifetime management of the aortic valve. This may include consideration for annular enlargement or root replacement in young patients with small annulus to accommodate appropriately sized valves for future reintervention.

Quality improvement extends beyond the boundaries of traditional heart surgery and these efforts must be directed in the care of all patients, including those undergoing TAVR procedures. A recent study by Desai and colleagues using joint STS and American College of Cardiology transcatheter valve therapy data highlighted 11% of TAVR programs were identified as providing care below the average level of performance. The collaborative STS/ American College of Cardiology Transcatheter Valve Therapy Registry has provided quality reporting and outcome benchmarking to improve patient care.

WEAPONIZATION OF QUALITY DATA

Human nature drives every surgeon and every program to perform the highest quality surgery and strive for superior outcomes in comparison to peers. Quality reporting may provide a measuring stick by which to compare performance between programs; however, this was not and is not the intended purpose of the data. To address these issues, the STS developed a participant-level star rating system to provide meaningful feedback to programs if they are below expected (1 star), as expected (2 stars), or above expected (3 stars) compared with the overall population. To provide the most unbiased system, the STS Quality Measurement Task Force selected a 98% Bayesian credible interval as its outlier criterion, corresponding to a 99% true Bayesian probability that a participant is a high or low outlier. Therefore, single center attempts at weaponizing their quality data to demonstrate superiority simply detracts from the overall purpose of quality measurement.

Johnston and colleagues overly strong conclusion about the inappropriateness of STS Risk Models for estimating contemporary SAVR risk was based on data from a single, large-volume institution. This implied that other large-volume centers are likely to show similar findings, whereas
smaller programs might not. Notably, these conclusions were based on the authors’ reported comparisons with risk estimates based on the 2008 STS Risk Models rather than the updated July 2017 models. I wonder if a similar analysis of the same institution’s 2017 to present SA VR data compared with current STS Risk Model estimates—and appropriate recalibration—would have yielded similar findings and conclusions. Even with the newer models, STS continues to provide recalibration factors updated 4 times per year (with how-to-use instructions) to participants, and these should be used to accurately estimate risk. Setting aside the known pitfalls of using a single institution’s comparison data to draw conclusions about the general accuracy and agility of the STS SAVR risk models derived from more than 1000 programs nationally, it should be noted that the authors did not apply the required calendar-year-based recalibration factor the STS provides to database participants to adjust expected SAVR mortality estimates such that the national yearly O:E SAVR mortality rates are balanced (ie, overall O:E = 1).

ACCESS TO CARDIAC SURGERY CARE

Several studies have supported the centralization of surgical intervention on cardiac valve disease to optimize outcomes through development of reference centers. Whereas centralization of care may provide superior outcomes, it is critical to ensure adequate access to these referral centers for patients with complex valve disease. Alternatively, it would be more beneficial to raise the quality of all programs through targeted initiatives, shared protocols and care pathways, as well as centralized data collection as supported by the STS. This approach to quality improvement and data reporting mitigates the unintended consequences of risk aversion and loss of access to life-saving aortic valve surgery.

CONCLUSIONS

The STS Risk Models are evolving with the field of cardiac surgery and provide real-world contemporary risk stratification in a user-friendly calculator tool. The volume-outcome relationship is well documented, but the
Conflict of Interest Statement

Dr Bavaria has financial relationships with Edwards Lifesciences, Medtronic, and Abbott Laboratories.

The Journal policy requires editors and reviewers to disclose conflicts of interest and to decline handling manuscripts for which they may have a conflict of interest. The editors and reviewers of this article have no conflicts of interest.

Dr Bavaria served as the 52nd President of the Society of Thoracic Surgeons.

References


Key Words: aortic valve replacement, quality outcomes, transcatheter aortic valve replacement, surgical risk