The Society of Thoracic Surgeons Mitral Valve Repair/Replacement Plus Coronary Artery Bypass Grafting Composite Score: A Report of The Society of Thoracic Surgeons Quality Measurement Task Force

J. Scott Rankin, MD, Vinay Badhwar, MD, Xia He, MS, Jeffrey P. Jacobs, MD, James S. Gammie, MD, Anthony P. Furnary, MD, Frank L. Fazzalari, MD, Jane Han, MSW, Sean M. O'Brien, PhD, and David M. Shahian, MD

Division of Cardiothoracic Surgery, West Virginia University, Morgantown, West Virginia; Duke Clinical Research Institute, Durham, North Carolina; Johns Hopkins All Children's Heart Institute, All Children's Hospital, Johns Hopkins University School of Medicine, Saint Petersburg, Florida, and Baltimore, Maryland; Department of Cardiac Surgery, University of Maryland, Baltimore, Maryland; Starr-Wood Cardiac Group, Portland, Oregon; Section of Cardiac Surgery, University of Michigan, Ann Arbor, Michigan; The Society of Thoracic Surgeons, Chicago, Illinois; and Department of Surgery, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts

Background. **The Society of Thoracic Surgeons (STS) Quality Measurement Task Force has developed a composite performance measure for mitral repair/replacement (MVRR) with concomitant coronary artery bypass grafting (CABG).**

Methods. **Data were acquired from the STS Adult Cardiac Surgery Database for 26,463 patients undergoing MVRR + CABG operations between July 1, 2011, and June 30, 2014. Established STS risk models were applied, along with modifications enabling the inclusion of patients with concomitant closures of atrial septal defects and patent foramen ovale, surgical ablation for atrial fibrillation, and tricuspid valve repair (TVR). Participants with fewer than 10 eligible cases over 3 years were excluded. The MVRR + CABG composite consisted of two domains: risk-adjusted mortality and the any-or-none occurrence of major morbidity (prolonged ventilation, deep sternal infection, permanent stroke, renal failure, and reoperation). Composite performance scores were calculated with the use of hierarchic regression models, and high-performing and low-performing outliers were determined with the use of 95% Bayesian credible intervals.**

Results. **There were 24,740 patients at 703 participant sites after exclusions. Two percent (14/703) of programs were classified as 1-star (lower than expected performance), 95% (666/703) were classified as 2-star (as-expected performance), and 3% (23/703) were classified as 3-star (higher than expected performance). The average unadjusted operative mortality was 6.2% (1,532/24,740), and a monotonic decline in both mortality and morbidity was observed as star rating scores increased.**

Conclusions. **An STS composite performance measure was developed for MVRR + CABG operations. This measure may be useful for outcome assessment, quality improvement, patient counseling, clinical research, and public reporting.**

(Ann Thorac Surg 2017;103:1475–81)

The Society of Thoracic Surgeons (STS) has developed operative risk models and composite performance measures for isolated coronary artery bypass grafting (CABG), isolated aortic valve replacement (AVR), and AVR + CABG [1–7]. The CABG composite measure consists of four domains: (1) risk-adjusted mortality, (2) risk-adjusted any-or-none major morbidity (renal failure, permanent stroke, reoperation, deep sternal infection, prolonged ventilation), (3) use of at least one internal mammary artery bypass graft, and (4) use of all perioperative medications endorsed by the National Quality Forum. The two AVR composite measures consist of only the first two of those domains because widely accepted process measures are not available. These STS composite

Dr Rankin discloses a financial relationship with BioStable Science and Engineering Inc., Admedus Corp, and AtriCure USA.

The Supplemental Tables can be viewed in the online version of this article [http://dx.doi.org/10.1016/j.athoracsur.2016.09.035] on http://www.annalsthoracicsurgery.org.

Accepted for publication Sept 7, 2016.

Address correspondence to Dr Rankin, Professor of Surgery, WVU Heart and Vascular Institute, Morgantown, WV; email: jsrankinmd@cs.com.

0003-4975/$36.00
http://dx.doi.org/10.1016/j.athoracsur.2016.09.035

measures have been useful for quality assessment, practice improvement, patient counseling, research, and public reporting.

A composite performance measure for isolated mitral valve repair or replacement (MVRR) recently was developed in a companion study [8]. A clinically related procedure, MVRR + CABG, constitutes an increasing proportion of cardiac surgical practice, and mortality risk is higher than for isolated MVRR [9–13]. An STS composite performance measure for MVRR + CABG has been developed to enable benchmark comparisons among STS participants and to facilitate outcome assessment and quality improvement.

Material and Methods

Patient Population

The study population consisted of 26,463 adult patients undergoing MVRR + CABG in North America between July 1, 2011, and June 30, 2014. Data were collected by use of the STS Adult Cardiac Surgery Database (ACSD) version 2.73, and all patients receiving MVRR + CABG were initially included. Patients who had arrhythmia devices (eg, internal cardiac defibrillators), transmyocardial revascularization, concomitant vascular or pulmonary procedures, prior mitral clip, and missing age, sex, or both were subsequently excluded, as were STS participants outside the United States or those with fewer than 10 eligible cases over 3 years. The study population included patients with any acuity status (including emergency and salvage), those with closure of atrial septal defects or patent foramen ovale, operations for endocarditis (active or treated), reoperations, surgical ablation procedures (both intracardiac and extracardiac) for atrial fibrillation (AF), and concomitant tricuspid valve repair (TVR). These inclusion and exclusion criteria differ slightly from the STS 2008 risk models [1–3] and were selected to better reflect evolving science and practice trends. For example, discretionary procedures such as concomitant TVR are usually not included in risk models. However, we did so in this instance for two reasons. First, TVR may serve as an additional marker beyond severity of tricuspid regurgitation for more advanced tricuspid disease and right ventricular dysfunction. Second, TVR may confer long-term benefits that outweigh some potential short-term risks, and we did not want to discourage TVR by failing to adjust for any potential impact on early risk. The final study population comprised 24,740 operations among 703 STS participating centers.

Estimation of Risk-Adjusted Outcome Measures

The composite measure is a weighted combination of a participant's risk-adjusted operative mortality (OM) and risk-adjusted major morbidity rates. Operative mortality was defined as death before hospital discharge or within 30 days of operation. Major morbidity (an any-or-none outcome) included postoperative prolonged ventilation, deep sternal infection, permanent stroke, renal failure, and reoperations. To adjust for case mix, logistic regression models for operative mortality and major adverse events were estimated by the use of covariates from published STS 2008 risk models [2, 3]. The etiologies of mitral valve disease were not included in the final model because of unacceptably high missing data rates (24.7%).

Each model's fit to the data was assessed by a comparison of observed versus expected outcomes within subgroups and across deciles of predicted risk. The subgroups were based on presence of a tricuspid procedure and amount of tricuspid insufficiency (none to mild, moderate, severe). After confirmation of satisfactory calibration, the models were used to calculate each participant's expected rates of OM and major adverse events. The expected rates then were entered as risk scores in a Bayesian hierarchical model that simultaneously estimated rates of OM and major morbidity for each participant.

Estimation of the Composite Measure Score and Star Ratings

Consistent with previous composite measures, risk-adjusted event rates first were converted into risk-adjusted absence-of-event rates. To calculate the composite score, participant-specific absence of mortality rates and absence of morbidity rates were weighted inversely by their respective standard deviations across participants. This procedure was equivalent to first rescaling the absence of mortality rates and absence of morbidity rates by their respective standard deviations across participants, and then assigning equal weighting to the rescaled rates. Finally, to draw statistical inferences about participant performance, a Bayesian credible interval surrounding each participant's composite score was calculated. Unlike frequentist confidence intervals, a Bayesian credible interval has an intuitively direct interpretation as an interval containing the true value of the composite score with a specified probability (eg, 95%).

To determine star ratings for each participant, the credible interval of its composite score was compared with the STS average. Participants whose intervals were entirely above the STS average were classified as 3-star (higher than expected performance), and participants whose intervals were entirely below the STS average were classified as 1-star (lower than expected performance). Credible intervals based on different probability levels (90%, 95%, 98%) were explored, and the resulting percentages of 1-star, 2-star, and 3-star programs were calculated.

The reliability of the composite score was estimated as the squared correlation between the calculated composite score and the true score as described previously [7]. Briefly, reliability may be interpreted as the proportion of variation in a measure that is attributable to true differences between the measured units (ie, signal) as opposed to random statistical fluctuations (ie, noise). As in previous STS composite measure development, our goal was to achieve as high a reliability as possible (at least 0.50), which generally required establishing a minimum number of procedures performed over a 3-year period for