
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.


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 measures include scores for superior, average, and substandard care.

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.annalsthoracic.org.
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
eligibility. This goal had to be balanced by the competing goal of providing a score to as many centers as possible.

**Sensitivity Analysis: Mitral Disease Etiology**
Etiology was not included as a covariate in the risk model, mainly because 24.7% of patients had etiology listed as “other” or “missing” (Supplemental Table A). The model development team hypothesized that other consistently collected risk variables in the model were the underlying factors leading to the apparent association of etiology with outcomes, as shown previously [10, 14, 15], and that the absence of a specific etiologic variable would not affect model performance. To further explore whether mitral disease etiology was an independent risk predictor (thus compromising a model that did not include it), we examined the degree to which outcome comparisons between STS participants and the national benchmark might be confounded by unadjusted differences in the mix of mitral disease etiologies. A sensitivity analysis was conducted with the records of patients with nonmissing etiologies (75.3%). An “augmented” operative mortality model was estimated by adding recorded mitral disease etiology as a categoric variable. We calculated participant-specific expected mortality rates and risk-adjusted OM rates by use of the final STS model and the augmented model and compared the MVRR + CABG results from these two models among participants having at least 30 eligible cases with nonmissing mitral disease etiologies.

**Sensitivity Analysis: Expanded Inclusion Criteria**
To evaluate the effect of the expanded patient inclusion criteria for this measure compared with prior STS mitral models, a sensitivity analysis was performed by excluding the “active endocarditis” patients from the model and by documenting the absolute differences in composite scores. Pearson and Spearman rank correlations then were estimated between the two sets of scores.

The Duke University Institutional Review Board granted a waiver of informed consent for use of these registry data for quality assurance.

**Results**
The baseline characteristics for the 24,740 patients are shown in Supplemental Table A. The median age was 69 years. The etiology of mitral disease was classified as degenerative in 51.8% (12,807/24,740), ischemic in 13.3% (3,300/24,740), rheumatic in 4.6% (1,129/24,740), endocarditis in 2.6% (643/24,740), and “other” or missing in 24.7% patients. Mitral repair was performed in 16,300 (65.9%) of patients, and 8,440 had valve replacement (34.1%), which was more common in the rheumatic and endocarditis categories. Concomitant TVR was performed in 2,712 (11.0%), and the incidence was 1.6% higher in the mitral replacement than in the mitral repair subgroups. The overall unadjusted OM was 6.2% (1,532/24,740), and one or more major morbidities occurred in 30.8% (7,620/24,740), with prolonged ventilation being the most common (26%) (Table 1). When prolonged ventilation was observed, the associated unadjusted OM increased from 2.4% to 17.1% (Supplemental Table B). The occurrence of any major morbidity was associated with an increased unadjusted OM to 15.7% of patients, compared with 2.0% without major adverse events. Thus, pulmonary complications proved to be the major driver of morbidity-associated mortality.

The distributions of adjusted mortality and morbidity and estimated composite scores across the 703 STS participant centers are presented in Figures 1 and 2. The weights of mortality and morbidity domains in the composite score calculation were 0.78 and 0.22, respectively. The numbers and percentages of 1-star, 2-star, and 3-star programs for various Bayesian credible intervals are displayed in Table 2, and a 95% credible interval was selected. By use of this criterion, 14 of 703 centers (2%) were assigned a 1-star rating (lower than expected performance), 23 of 703 (3%) were 3-star (better than expected performance), and the remaining 666 (95%) were 2-star (as-expected performance).

The aggregate adjusted OM and morbidity rates by star rating category are shown in Table 3. In comparison with centers in the “as expected” 2-star category, adjusted OM approximately doubled and halved for the 1-star and 3-star categories, respectively. The morbidity rate differences across star rating categories were directionally similar but slightly less in magnitude. These findings provided internal validation that the performance score measured what it purported: the overall quality of MVRR + CABG.

The overall composite measure reliability was greater than 0.50 in the 341 centers performing 25 cases or more over 3 years. Thus, about half of the total number of programs would be eligible to receive a star rating (Supplemental Table C). Higher reliability could be achieved by further substantial reductions in the number of eligible programs, but that was thought to be counterproductive.

As shown in Figure 3, the adjusted OM rates calculated with and without adjustment for etiology of mitral disease were nearly identical, with a Pearson correlation coefficient of 0.9985. The ratio of risk-adjusted OM rates

---

**Table 1. Number of Participants, Operations, and Events**

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time duration of data set</td>
<td>3 years</td>
</tr>
<tr>
<td>Participants, n</td>
<td>703</td>
</tr>
<tr>
<td>Operations, n</td>
<td>24,740</td>
</tr>
<tr>
<td>Mortality, n (%)</td>
<td>1,532 (6.2)</td>
</tr>
<tr>
<td>Any major morbidity, n (%)</td>
<td>7,620 (30.8)</td>
</tr>
<tr>
<td>Prolonged ventilation</td>
<td>6,428 (26.0)</td>
</tr>
<tr>
<td>Deep sternal infection</td>
<td>102 (0.4)</td>
</tr>
<tr>
<td>Permanent stroke</td>
<td>684 (2.8)</td>
</tr>
<tr>
<td>Renal failure</td>
<td>1,601 (6.5)</td>
</tr>
<tr>
<td>Reoperations⁶</td>
<td>1,350 (5.5)</td>
</tr>
</tbody>
</table>

⁶ Reoperation (1) for bleeding, (2) for intervention of coronary graft occlusion because of acute closure, thrombosis, technical, or embolic origin, (3) for prosthetic or native valve dysfunction, and (4) for other cardiac reasons.
calculated with versus without adjustment for etiology of mitral disease ranged from 0.934 to 1.140 (interquartile range [IQR] 0.988 to 1.008) in the MVRR + CABG population. On the absolute scale, the difference in risk-adjusted OM estimates based on the final versus augmented model was always less than 0.1 of the width of the 95% confidence interval.

In the sensitivity analysis for inclusion of active endocarditis cases, the average absolute difference between the original composite score and the composite score without 416 active endocarditis patients was 0.0015 (IQR 0.0008 to 0.0018). On the relative scale, the median relative change from the original to the new score was 0.05% (IQR 0.03% to 1.13% (relative change was defined as new score / original score). The Pearson correlation between the two sets of scores was 0.995, and the Spearman rank correlation was 0.993.

Comment

The STS performance measures have been developed from the STS ACSD, a clinical registry with nearly 95% national penetration among adult cardiac surgery centers. Through annual external audits of 10% of participant

<table>
<thead>
<tr>
<th>CrI</th>
<th>1 star, n (%)</th>
<th>2 star, n (%)</th>
<th>3 star, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>90% CrI</td>
<td>23 (3.3)</td>
<td>647 (92.0)</td>
<td>33 (4.7)</td>
</tr>
<tr>
<td>95% CrI</td>
<td>14 (2.0)</td>
<td>666 (94.7)</td>
<td>23 (3.3)</td>
</tr>
<tr>
<td>98% CrI</td>
<td>9 (1.3)</td>
<td>682 (97.0)</td>
<td>12 (1.7)</td>
</tr>
</tbody>
</table>

Mortality domain (N = 703)

<table>
<thead>
<tr>
<th>CrI</th>
<th>1 star, n (%)</th>
<th>2 star, n (%)</th>
<th>3 star, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>90% CrI</td>
<td>1 (0.1)</td>
<td>700 (99.6)</td>
<td>2 (0.3)</td>
</tr>
<tr>
<td>95% CrI</td>
<td>0 (0.0)</td>
<td>702 (99.9)</td>
<td>1 (0.1)</td>
</tr>
<tr>
<td>98% CrI</td>
<td>0 (0.0)</td>
<td>702 (99.9)</td>
<td>1 (0.1)</td>
</tr>
</tbody>
</table>

Morbidity domain (N = 703)

<table>
<thead>
<tr>
<th>CrI</th>
<th>1 star, n (%)</th>
<th>2 star, n (%)</th>
<th>3 star, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>90% CrI</td>
<td>24 (3.4)</td>
<td>648 (92.2)</td>
<td>31 (4.4)</td>
</tr>
<tr>
<td>95% CrI</td>
<td>14 (2.0)</td>
<td>672 (95.6)</td>
<td>17 (2.4)</td>
</tr>
<tr>
<td>98% CrI</td>
<td>7 (1.0)</td>
<td>687 (97.7)</td>
<td>9 (1.3)</td>
</tr>
</tbody>
</table>

CrI = Bayesian credible intervals.

<table>
<thead>
<tr>
<th>Rating</th>
<th>1-star</th>
<th>2-star</th>
<th>3-star</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mortality (%)</td>
<td>11.6</td>
<td>6.3</td>
<td>3.5</td>
</tr>
<tr>
<td>Morbidity (%)</td>
<td>52.6</td>
<td>31.1</td>
<td>21.0</td>
</tr>
</tbody>
</table>

Risk-adjusted

<table>
<thead>
<tr>
<th>CrI</th>
<th>1-star</th>
<th>2-star</th>
<th>3-star</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mortality (%)</td>
<td>11.2</td>
<td>6.1</td>
<td>3.3</td>
</tr>
<tr>
<td>Morbidity (%)</td>
<td>51.0</td>
<td>31.0</td>
<td>20.3</td>
</tr>
</tbody>
</table>

CrI = Bayesian credible interval.
programs, the STS ACSD also has been shown to contain highly accurate data (96% to 97% congruence with medical record abstractions). On the basis of these findings and with the use of appropriate statistical modeling, an extensive portfolio of risk models and performance measures for the outcomes of five specific operations have been developed: isolated CABG, isolated AVR, AVR + CABG, isolated MVRR, and now MVRR + CABG. Importantly, the conceptual and technical details of these models and measures are fully transparent and have been published in their entirety in the peer-reviewed literature [1–8].

Compared with single-outcome metrics, such as mortality, composite performance measures have proved to be even more valuable tools for contemporary outcome assessment and quality improvement [4–7]. The advantages of composite measures include higher effective sample sizes / event rates compared with single outcomes, and more comprehensive evaluation of performance than could be achieved with the use of mortality alone. Finally, the inclusion of higher-risk categories, such as active endocarditis, made the current analysis more comprehensive but did not appreciably influence results.

The findings of this study provide interesting clinical insights into what has been one of the historically higher-mortality procedures in cardiac surgery [9]. Outcomes have improved in recent years as a result of innovations in surgical technique and patient care. For example, the data in the current study revealed contemporary unadjusted OM of 4.9% for mitral repair plus CABG and 8.7% for mitral replacement plus CABG (6.2% overall) (Supplemental Table 1). Although comparative analysis is difficult, it is likely that the transition of two thirds of MVRR + CABG patients from replacement to lower-risk mitral repair is a major factor responsible for the overall outcome improvement [9–13]. Better processes of care also have contributed, including augmented myocardial protection, improved cardiopulmonary bypass, and numerous innovations in intraoperative and postoperative management. In the present study, the degree to which pulmonary adverse events contributed to increased mortality and morbidity in the MVRR + CABG population was impressive. From previous STS ACSD work, pulmonary complications seem to vary widely between centers [16] and are increasing over time [17]. Thus, better processes of care in this one area could improve the overall results significantly [18, 19], and they represent an opportunity for near-term quality improvement.

The current 65.9% mitral repair rate for MVRR + CABG patients is increased over that observed a decade ago [20], but the denominator of this rate also includes patients undergoing reoperation who may not be candidates for repair. If reoperations after previous mitral replacement were removed, then the primary mitral repair rate approximated 75% of candidates and represents a substantive increase. A similar repair rate of greater than 70% was observed for isolated MVRR in a companion study derived from the same 2.73 data set [8]. With better early and late results uniformly observed after mitral repair [9–13], conversion to predominant valve reconstruction should be encouraged.

Several other findings also merit discussion. Half of the MVRR + CABG patients had degenerative causes of disease. Ninety-five percent of repair patients had annuloplasty, 16% had leaflet resection, 8% had artificial chordal replacement, and only 13% had leaflet augmentation patches. The increasing application of adjunctive repair techniques may further expand the proportion of cases resulting in effective and durable repair [21]. Interestingly, the baseline characteristics and procedural incidences in the MVRR + CABG population differed little from isolated mitral surgical procedures [8]. However, mortalities and morbidities were higher with the addition of ischemic heart disease to the pathophysiology.

The results of concomitant TVR appear to be improving. In previous reports, performance of a tricuspid procedure independently increased short-term risk [22–25], although late outcomes probably were better. In a separate analysis of the current data set [26], short-term mortality for mitral procedures with concomitant TVR was not statistically significantly higher, regardless of the degree of tricuspid regurgitation [26]. This finding is possibly due to advancing surgical techniques and to better operative and postoperative management. Improving results also may reflect increasing experience with TVR, now approaching 95% of all tricuspid cases [25]. Conversion to more effective tricuspid annuloplasty, such as geometric rings, may have contributed [27, 28]. Surgical ablation for AF has not been associated with higher early mortality [29], as also was the case in the current series. In fact, effective surgical ablation in patients with preoperative AF currently may be protective [30]. This finding would support more liberal...
application of concomitant ablation procedures, although long-term data are still required.

Finally, the nearly perfect correlation between adjusted OM rates with the “augmented” model (including the recorded variable in etiology of disease) and adjusted OM rates without etiology suggested that no significant information was added by including etiology, consistent with previous studies [10, 14, 15]. Thus, the apparent influence of etiology of ischemic disease on OM seem mediated by the effects of other included risk factors, rather than assignment of etiology per se. Similarly, the inclusion of high-risk categories, such as endocarditis, in the model did not diminish predictive accuracy and could even improve discrimination by increasing events.

Limitations
The combination of high center penetrance [31], a 10% annual center audit rate, large sample sizes, and linkages to other databases [32] have made the STS ACSD a valuable resource for the study of cardiothoracic procedures. However, it must be acknowledged that because participation in the STS ACSD remains voluntary, the results could be skewed toward better-performing centers. This seems unlikely, given the high national participation in the STS ACSD (90% to 95%).

Data from 703 sites were used to develop the MVRR + CABG measure. However, to achieve a reliability of 0.50, it was necessary to require a minimum volume of 25 cases over 3 years for centers to be eligible to receive composite scores, and this threshold was met by only 341 programs. The proportion of participant centers eligible to receive a MVRR + CABG composite score was smaller than for previous STS composite measures because of the generally lower case volumes per center for this procedure. For example, the total number of MVRR + CABG procedures available for model development was 24,740, compared with CABG (774,881), AVR (67,138), AVR + CABG (53,827), and isolated MVRR (61,201). By requiring a minimum volume threshold, we assure that those programs receiving an STS MVRR + CABG score can have confidence in its reliability.

Conclusion
In summary, an STS composite performance measure incorporating adjusted OM and morbidity has been developed for MVRR + CABG operations. Based on 3-year STS data samples and 95% Bayesian credible intervals, 2% of STS ACSD participants had worse than expected performance, 95% had “as expected” performance, and 3% had better than expected performance. STS composite measures may be useful for outcome assessment, quality improvement, public reporting, and future clinical investigation.

References


