

The STS Mitral Repair/Replacement Composite Score: A Report of the STS Quality Measurement Task Force

¹Vinay Badhwar, MD, ²J. Scott Rankin, MD, ³Xia He, MS, ⁴Jeffrey P. Jacobs, MD, ⁵James S. Gammie, MD, ⁶Anthony P. Furnary, MD, ⁷Frank L. Fazzalari, MD, ⁸Jane Han, MSW, ³Sean M. O'Brien, PhD, and ⁹David M. Shahian, MD

¹University of Pittsburgh, Pittsburgh, Pennsylvania;

²Cardiothoracic Surgery Associates, Nashville, Tennessee;

³Duke Clinical Research Institute, Durham, North Carolina;

⁴Johns Hopkins University School of Medicine, Saint Petersburg, Florida, and Baltimore, Maryland;

⁵University of Maryland, Baltimore, Maryland;

⁶Starr-Wood Cardiac Group, Portland, Oregon;

⁷University of Michigan, Ann Arbor, Michigan;

⁸The Society of Thoracic Surgeons, Chicago, Illinois

⁹Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts

Running head: Mitral Valve Surgery Composite Measure

Keywords: Mitral valve repair, Mitral valve replacement, Outcomes, Modeling

Conflicts of Interest: None

Word Count: 4,056

Abstract: 220

Address Correspondence to:

Vinay Badhwar, MD

Department of Cardiothoracic Surgery

University of Pittsburgh

200 Lothrop Street, C-721,

Pittsburgh, PA, 15213

e-mail: badhwarv@upmc.edu

Abstract

Background. The Society of Thoracic Surgeons (STS) Quality Measurement Task Force is developing a portfolio of composite performance measures for the most commonly performed procedures in adult cardiac surgery. We now describe the fourth in this series, the STS composite measure for mitral valve repair/replacement (MVRR).

Methods. We examined all patients undergoing isolated MVRR with or without concomitant performance of tricuspid valve repair (TVr), surgical arrhythmia ablation, or repair of atrial septal defect (ASD) between July 1, 2011 and June 30, 2014. In this two-domain model, risk-adjusted mortality and any-or-none major morbidity were combined into a composite score using 3 years of STS data and 95% Bayesian credible intervals to calculate composite scores and star ratings.

Results. There were 61,201 MVRR patients studied at 867 participant sites. Mitral valve repair (MVr) was performed in 57.4% (35,114/61,201) and replacement (MVR) in 42.6% (26,087/61,201). Mortality was 2.9% (1,773/61,201) and occurrence of any major morbidity was 17.0% (10,381/61,201). The median composite score was 93.2% (interquartile range, 92.3% to 94.2%). Star rating classifications included 23/867 (2.6%) one-star programs (lower than expected performance), 795/867 (91.7%) two-star (as-expected or average performance) programs, and 49/867 (5.7%) three-star programs (higher than expected performance).

Conclusions. STS has developed a MVRR composite performance measure that will be used for participant feedback, quality performance assessment and improvement, and voluntary public reporting.

Clinical outcomes are the ultimate measures of healthcare performance, and they unify the interests of all stakeholders [1]. As outcome measures have become the preferred approach to assess quality in cardiothoracic surgery, The Society of Thoracic Surgeons (STS) has emphasized the importance of accurate clinical data, homogeneous target populations, robust risk adjustment, multi-dimensional endpoints, state of the art statistical methodologies, and appropriate methods for outlier determination [2-4].

Beginning in 2007, the STS Quality Measurement Task Force has developed a series of procedure-specific composite performance measures based on risk-adjusted, homogeneous patient cohorts from the STS Adult Cardiac Surgery Database (ACSD) [3-6]. These STS procedure-based composite measures have received National Quality Forum (NQF) endorsement and provide the basis for STS public reporting initiatives utilizing star-rating categories ranging from 1-star (lowest performance) to 3-star (highest performance).

The STS coronary artery bypass grafting (CABG) performance measure was the first STS composite. It comprised 11 NQF-endorsed measures and consisted of four domains – risk-adjusted mortality, risk-adjusted occurrence of any of five major complications (prolonged ventilation > 24hrs, deep sternal wound infection, reoperation, renal failure or stroke), use of at least one internal mammary artery bypass graft, and use of all four NQF-endorsed perioperative medications [3,4]. Given the large number of CABG procedures, it was possible to apply a 99% Bayesian probability criterion to classify high or low performance groups based on 1-year data. STS subsequently developed similar composite measures for isolated aortic valve replacement (AVR) and AVR+CABG [5,6]. As no generally accepted or NQF-endorsed process measures existed for valve procedures, these valve composites consisted of two rather than four domains – risk-adjusted mortality and morbidity. Because these operations are generally not performed at the frequency of CABG at most STS participating sites, 1-year data were not felt to provide

adequate sample size for these measures. For this same reason, it was not appropriate to require 99% Bayesian probability, which corresponds to 98% Bayesian Credible Intervals (CrI). With the smaller number of cases, requiring 99% Bayesian probability would result in almost all programs classified as average, with very few high or low performing participants. Given these considerations, 3 years of data and 95% CrI were used for the final implementation.

Mitral valve operations are being performed with increasing frequency for a variety of etiologies and pathologies [7-9]. Building upon our three prior composite performance models, this report describes the development and testing of the STS composite performance measure for isolated mitral valve repair/replacement (MVRR), with or without concomitant tricuspid valve repair (TVr), surgical ablation for atrial fibrillation (AF), or repair of atrial septal defect (ASD).

Material and Methods

Population Studied

There were a total of 120,205 mitral valve operations from 1,082 centers in the STS ACSD from July 1, 2011 to June 30, 2014 (3 years). We identified 62,545 patients from 1,070 centers who underwent isolated MVRR operations with or without TVr, surgical ablation for AF, or ASD closure. After exclusions for prior MitraClip™ (Abbot Vascular, Abbot Park, IL), participants outside the US, and those sites with fewer than 10 eligible cases over 3 years, our study population consisted of 61,201 patients from 867 centers (Figure 1). Any patients with additional planned concomitant operations other than those noted were excluded. Any unplanned operations due to iatrogenic surgical complications were included in the analysis.

Given similar sample size issues to those encountered with our isolated AVR and AVR+CABG models [5,6], we performed exploratory analyses using various credible intervals to determine those most appropriate for this measure.

Estimation of Risk-Standardized Outcome Measures

To adjust for case mix, the published 2008 STS isolated valve model [10] was utilized but modified, consistent with evolving trends in mitral surgery. Given the frequent performance of concomitant ASD closure and surgical ablation procedures for AF, the incremental risks associated with each procedure were investigated within this dataset. As no significant incremental risk increases were found to be associated with concomitant ASD closure or AF ablation, patients who had these concomitant procedures were included in the analysis.

Conversely, exploratory analyses of concomitant TVr performance revealed that this increased the morbidity of mitral surgery. Given the frequent concomitant performance of TVr, this was added to the measure and included as a risk factor to the models. This modeling decision merits explanation. It is a generally accepted principle not to use what may be discretionary procedural decisions (e.g., whether or not to add a tricuspid valve repair) in profiling models. However, there is accumulating evidence of the potential longitudinal merits of concomitant TVr [12,13], and it was our goal not to discourage the performance of this procedure by failing to account for its increased inherent risk of morbidity. Furthermore, it is likely that the need to perform TVr may be a proxy for more advanced disease, such as subclinical right ventricular dysfunction, that may not be accounted for by measuring only the degree of tricuspid regurgitation (TR).

We also modified existing mitral valve models to more precisely quantify the degree of TR from two categories – none-mild and moderate-to-severe – to none-mild, moderate, or

severe in the new measure. Finally, we expanded and clarified endocarditis categories from only two – active or other – to three: active, treated, or none.

All variables included in the models had less than 1% missing data except ejection fraction (2.5%). Missing values were handled using the same method as that used by the 2008 risk model [10]. Missing data on binary (yes/no) and categorical variables were imputed to the lowest risk value, whereas missing data on continuous variables were imputed to group-specific medians.

Using the modified models, we calculated participant-specific average expected event rates, which were then entered as covariates in a Bayesian hierarchical model that estimated risk-standardized mortality and morbidity rates simultaneously. Methods of estimating these quantities were congruent with previously published STS composites [3-6].

Estimation of Composite Scores and Star Ratings

The statistical methodology used to estimate the STS MVRR composite score and star rating for each participant site was similar to that used for the STS isolated CABG, isolated AVR, and AVR+CABG measures [3-6]. As with previous composite scores, we first translated risk-standardized event rates into risk-standardized absence of event rates so that a higher score indicated better performance. We then rescaled the morbidity and mortality domains by dividing by their respective standard deviations, then added the two domains together.

Reliability Estimation

As in our previous valve measures [5,6], we estimated reliability as the squared correlation between the calculated MVRR composite score and the true score, the latter estimated using Markov Chain Monte Carlo simulations (Appendix A).

Results

In the final study cohort of 61,201 patients, 35,912 (58.7%) were isolated MVRR operations and 25,289 (41.3%) had at least one of the included concomitant procedures. Concomitant TVr was performed in 15.8% (9,698/61,201); 29.3% (17,922/61,201) had surgical ablation; and 6.4% (3,940/61,201) had ASD closure.

Mitral valve repair (MVr) was performed in 57.4% (35,114/61,201) of patients and mitral valve replacement (MVR) in 42.6% (26,087/61,201). Conventional sternotomy was the surgical approach in 72.5% (44,361/61,201) of the operations, and right thoracotomy was used in 14.2% (8,668/61,201). Robotic assistance was applied in 7.0% (4,278/61,201), more often for MVr procedures, 11.2% (3,948/35,114). AF was present in 32.2% (19,689/61,201) and a surgical ablation procedure was performed in 61.5% (12,102/19,689) of these patients.

Mitral valve pathology and etiology were documented in 87.5% (53,535/61,201) of the cohort. Of these, 56.5% (30,222/53,535) were identified as having mitral regurgitation (MR) due to annular or degenerative disease, without stenosis. MVr was performed in 75.0% (22,662/30,222) of these patients.

Table 1 reveals the numbers of participant sites, MVRR operations, and descriptive analyses of 3-year outcomes. Overall mortality was 2.9% (1,773/61,201) and the frequency of any major morbidity was 17.0% (10,381/61,201).

Figure 2 provides the estimated distributions of risk-standardized mortality and any-or-none morbidity across STS ACSD participant sites, both of which show substantial variation.

Figure 3 plots the distribution of the calculated MVRR composite scores; median score was 93.2% with an interquartile range of 92.3% to 94.2% across all participants. After rescaling,

the relative weights in the final composite of risk-standardized mortality and risk-standardized major morbidity were 0.74 and 0.26, respectively.

Table 2 shows the number and percentage of various star rating categories using 90%, 95%, and 98% Bayesian CrI. As in our previous AVR and AVR+CABG models, we determined that 95% CrI (corresponding to 97.5% Bayesian probability) was the most appropriate for the MVRR composite, as it provided the best compromise between sensitivity to identify outliers and specificity to minimize the possibility of false positive outlier identification. This resulted in 2.6% (23/867) of programs assigned 1-star, 91.7% (795/867) 2-stars, and 5.7% (49/867) awarded 3-stars.

Table 3 shows the monotonic decrease in observed and risk-adjusted mortality and major morbidity as star ratings increase, regardless of the Bayesian CrI. Using 95% CrI, the 1-star, 2-star and 3-star MVRR risk-adjusted mortality rates were 6.8%, 3.1% and 1.2%, respectively, and risk-adjusted morbidity rates were 31.2%, 17.7% and 11.4%, respectively. These findings provide internal validation that the STS MVRR star ratings perform as desired, assigning higher scores to programs with lower rates of adverse events.

Figure 4 depicts the distribution of star rating by site procedural volume. This scatter plot illustrates that while 1-star performing centers appear at all levels of volume, those awarded 3-stars were generally higher volume centers.

Concomitant TVr at the time of MVRR operations was uncommon (15.8%; 9,698/61,201). Of the 867 participant sites analyzed, 190 performed TVr in at least 10% of their MVRR operations, fewer than 50 sites performed TVr between 25-50% of the time, and less than 10 sites performed concomitant TVr in greater than 50% of their cases.

The estimated reliability of the STS MVRR composite measure using 3 years of data in participants with at least 36 total cases was 0.58 (95% CrI, 0.52 to 0.64) as outlined in Table 4.

For comparison, reliability of the STS isolated CABG measure was 0.77 (95% CrI, 0.74 to 0.80) using 1 year of data in 2013. Using 3 years of data from 2011 to 2013, the reliability of the STS AVR composite measure was 0.52 (95% CrI, 0.47 to 0.57) and the AVR+CABG measure was 0.50 (95% CrI, 0.45 to 0.54) [5,6].

Comment

We have described the development and operating characteristics of the STS MVRR composite performance measure. This is based on a 3-year assessment of risk-adjusted mortality and risk-adjusted any-or-none major morbidity for isolated MVr or MVR with or without concomitant TVr, surgical arrhythmia ablation, or ASD repair.

Procedural Inclusiveness

The clinical presentation of mitral valve disease is quite different than that of aortic valve disease, in that patients often present with secondary structural or electrophysiological abnormalities that also require correction. Nearly one half of all isolated mitral operations in our cohort from the STS ACSD included at least one of the following procedures: a tricuspid valve repair, surgical ablation procedure and/or closure of patent foramen ovale or secundum ASD. Given that these concomitant procedures are the three most commonly associated with mitral operations both in the STS Adult Cardiac Surgery Database and in the peer-reviewed literature [7-9], we elected to adapt the 2008 model for isolated valve surgery [10] in order to include each of these concomitant procedures. The degree of tricuspid regurgitation in prior STS risk models was categorized as none/mild or moderate-to-severe. Because the degree of TR may impact risk, we categorized TR severity into none/mild, moderate or severe.

We found that there was no incremental risk associated with the concomitant performance of ASD repair or surgical ablation. Similarly, there was no incremental mortality risk associated with TVr but there was additional morbidity risk at each incremental degree of TR (Appendix B). Adding a concomitant TVr to a mitral operation was associated with an odds ratio of 1.36 for morbidity but no incremental risk of mortality (OR 0.98). While there is controversy surrounding the indications for and value of correcting TR by adding TVr to mitral operations, there is growing interest in pursuing TVr when the echocardiographic measurement of the tricuspid annulus is ≥ 40 mm, even when the degree of TR is mild-moderate [12,13]. Our analysis reveals that some US centers appear to adhere to this philosophy by performing TVr in a majority of their cases, whilst the majority performs concomitant TVr in less than half of their mitral operations. Given the incremental morbidity risk of TVr in none-mild TR, further longitudinal analyses of these data are required to determine the merit of adding a TVr to mitral operations.

Frequency of Mitral Valve Repair as a Potential Quality Metric

The objective of the current analysis was to develop a multi-dimensional performance measure for mitral operations and to use this to assess the performance of STS participant sites. Surgery remains the optimal therapy for the management of primary MR, and MVr has superior outcomes compared to MVR among all age ranges [14-17]. In the current cohort, 75% of patients with primary MR due to degenerative disease underwent MVr. We carefully considered whether it was possible to include the site-specific percentage of patients undergoing MVr (as opposed to MVR) for degenerative mitral regurgitation as a process domain in the composite measure. However, among 867 centers eligible for this study, the data field for mitral disease etiology was missing more than 5% of the time in 385 sites, more than 10% in 294 sites and more than 25% in 169 sites. Overall, nearly 20% of sites had missing etiology for more than one quarter of their cases. This would make it difficult to separate procedures with a

pure degenerative etiology from those with a functional or ischemic etiology where the superiority of MVr is less well-established. Therefore, we ultimately concluded that due to the high rate of missing data for etiology, we could not include rate of MVr for degenerative disease as a process domain within the composite measure at this time. We will continue to encourage center-level completion of the etiology field and we will re-evaluate this possibility in the future.

Limitations

Participation in the STS ACSD remains voluntary. Nevertheless, 90-95% of all programs performing adult cardiac surgery in the United States are participants, and the data are highly representative of overall practice in this country. Information on detailed mitral pathoetiology was not consistently available and thus its role in outcome performance of operations, such as repair rates for primary MR, was not feasible to include in the risk model. Our analyses determined that it was necessary to require 36 cases over three years as a minimum threshold for receiving a site-specific STS MVRR composite score, which enabled a reliability of 0.58 but reduced the number of programs eligible to receive a score from 867 to 462. A higher volume threshold would have yielded even higher reliability, but at the cost of further reducing the number of programs eligible to receive a score.

Conclusions

The STS has developed a composite performance measure for MVRR based on risk-adjusted mortality and risk-adjusted occurrence of any major morbidity. Results using this MVRR composite score will be provided to the STS ACSD in 2015, with voluntary public reporting of the measure commencing in 2016. A companion measure for MVRR + CABG has also been developed and will be provided concurrently.

REFERENCES

1. Porter ME. What is value in health care? *N Engl J Med* 2010; 363:2477-2481.
2. Krumholz HM, Brindis RG, Brush JE, et al. Standards for statistical models used for public reporting of health outcomes: an American Heart Association Scientific Statement from the Quality of Care and Outcomes Research Interdisciplinary Writing Group. *Circulation* 2006; 113:456-462.
3. Shahian DM, Edwards FH, Ferraris VA, et al; Society of Thoracic Surgeons Quality Measurement Task Force. Quality measurement in adult cardiac surgery: part 1 conceptual framework and measure selection. *Ann Thorac Surg* 2007;83(4 Suppl):S3–12.
4. O'Brien SM, Shahian DM, DeLong ER, et al. Quality measurement in adult cardiac surgery: part 2—statistical considerations in composite measure scoring and provider rating. *Ann Thorac Surg* 2007;83(4 Suppl):S13–26.
5. Shahian DM, He X, Jacobs JP, et al. The Society of Thoracic Surgeons Isolated Aortic Valve Replacement (AVR) Composite Score: a report of the STS Quality Measurement Task Force. *Ann Thorac Surg* 2012;94:2166–71.
6. Shahian DM, He X, Jacobs JP, et al. The STS AVR+CABG composite score: a report of the STS Quality Measurement Task Force. *Ann Thorac Surg* 2014;97:1604-9.
7. Gammie JS, Sheng S, Griffith BP, et al. Trends in mitral valve surgery in the United States: results from the Society of Thoracic Surgeons Adult Cardiac Database. *Ann Thorac Surg* 2009;87;1431–39.
8. Bolling SF, Li S, O'Brien SM, et al. Predictors of mitral valve repair: clinical and surgeon factors. *Ann Thorac Surg* 2010;90:1904–11.
9. Gammie JS, O'Brien SM, Griffith BP, Ferguson TB, Peterson ED. Influence of hospital procedural volume on care process and mortality for patients undergoing elective surgery for mitral regurgitation. *Circulation* 2007;115;881–86.
10. O'Brien SM, Shahian DM, Filardo G, et al. The Society of Thoracic Surgeons 2008 cardiac surgery risk models: part 2 – isolated valve surgery. *Ann Thorac Surg* 2009;88(1 Suppl):S23-42.
11. Badhwar V, He M, Rankin JS, et al. Performing concomitant tricuspid valve repair at the time of primary mitral operation is not associated with increased mortality. *J Thorac Cardiovasc Surg* 2015 (*Submitted*).
12. Dreyfus GD, Corbi PJ, Chan KM, Bahrami T. Secondary tricuspid regurgitation or dilatation: which should be the criteria for surgical repair? *Ann Thorac Surg* 2005;79:127-32.
13. Vahanian A, Alfieri O, Andreotti F, et al. Guidelines on the Management of Valvular Heart Disease (Version 2012). *Eur Heart J* 2012;33:2451-2496.
14. Enriquez-Sarano M, Akins CW, Vahanian A. Mitral regurgitation. *Lancet* 2009;18;373:1382-94.
15. Badhwar V, Peterson ED, Jacobs JP, et al. Longitudinal outcome of isolated mitral repair in older patients: results from 14,604 procedures performed from 1991 to 2007. *Ann Thorac Surg* 2012;94:1870-79.

16. Suri RM, Vanoverschelde JL, Grigioni F, et al. Association between early surgical intervention vs watchful waiting and outcomes for mitral regurgitation due to flail mitral valve leaflets. *JAMA* 2013;310:609-616.
17. Daneshmand MA, Milano CA, Rankin JS, et al. Influence of patient age on procedural selection in mitral valve surgery. *Ann Thorac Surg* 2010;90:1479-1486.

ABBREVIATIONS LEGEND

ACSD – Adult Cardiac Surgery Database

ASD – Atrial Septal Defect

AF – Atrial Fibrillation

CABG – Coronary Artery Bypass Grafting

CrI – Credible Interval

MVr – Mitral Valve Repair

MVR – Mitral Valve Replacement

MVRR – Mitral Valve Repair/Replacement

NQF – National Quality Forum

STS – The Society of Thoracic Surgeons

TVr – Tricuspid Valve Repair

TABLES

Table 1. *Sample Size and Outcomes*

Time	3-Year
Participants, n	867
Operations, n	61,201
Mortality, n (%)	1,773 (2.9%)
Any Major Morbidity*, n (%)	10,381 (17.0%)
Prolonged ventilation	8,002 (13.1%)
Deep sternal infection	89 (0.1%)
Permanent stroke	946 (1.5%)
Renal failure	1,730 (2.8%)
Reoperations	2,785 (4.6%)

*Any-or-none occurrence of morbidity defined as prolonged ventilation > 24 hours, deep sternal wound infection, new postoperative permanent stroke, new postoperative renal failure, reoperation (for bleeding, prosthetic or native valve dysfunction, other cardiac reasons).

Table 2. *Star Ratings for Various Credible Intervals*

	Composite		
	1-star	2-star	3-star
90% CrI	32	767	68
95% CrI	23	795	49
98% CrI	14	817	36
	Mortality		
	1-star	2-star	3-star
90% CrI	3	850	14
95% CrI	0	860	7
98% CrI	0	864	3
	Morbidity		
	1-star	2-star	3-star
90% CrI	29	795	43
95% CrI	15	824	28
98% CrI	8	839	20

CrI = Credible Interval.

Table 3. *Observed and Risk-adjusted Mortality and Morbidity Rates by Composite Score Star Rating*

Rate	Mortality, %			Morbidity, %		
	1-star	2-star	3-star	1-star	2-star	3-star
Observed						
90% CrI	5.9	3.2	0.6	24.4	13.7	7.6
95% CrI	6.4	3.2	0.6	25.3	13.6	7.6
98% CrI	5.8	3.1	0.5	27.3	13.4	7.5
Risk-adjusted						
90% CrI	6.4	3.2	0.6	24.8	13.5	7.8
95% CrI	6.8	3.1	0.6	25.9	13.4	7.7
98% CrI	6.6	3.1	0.5	26.2	13.2	7.8

Table 4. *Reliability of the Mitral Valve Composite Measure*

Time Span	Number of Participants Included	Number of Patients Included	Reliability $\hat{\rho}^2$ (95% PrI)
3 years	867	61201	0.47 (0.42, 0.52)
3 years, participants with at least 25 cases	597	56799	0.55 (0.49, 0.60)
3 years, participants with at least 36 cases	462	52841	0.58 (0.52, 0.64)
3 years, participants with at least 50 cases	349	48076	0.61 (0.54, 0.66)
3 years, participants with at least 100 cases	165	35198	0.69 (0.62, 0.76)

Reliability is conventionally defined as the proportion of variation in a performance measure that is due to true between-hospital differences (i.e., signal) as opposed to random statistical fluctuations (i.e., noise). A mathematically equivalent definition is the squared correlation between a measurement and the true value.

FIGURES

Figure 1 Consort flow diagram of cohort formation. Participant sites or centers are defined as an institution or practice performing adult cardiac surgery in North America that participates in the STD Adult Cardiac Surgery Database.

Figure 2 Estimated distribution of risk-adjusted mortality and morbidity rates (IQR = interquartile range)

Figure 3 Distribution of estimated Society of Thoracic Surgeons Mitral Valve Repair/Replacement surgery composite scores (IQR = interquartile range)

Figure 4 Scatter plot of composite score star rating by volume of eligible cases.

Figure 1

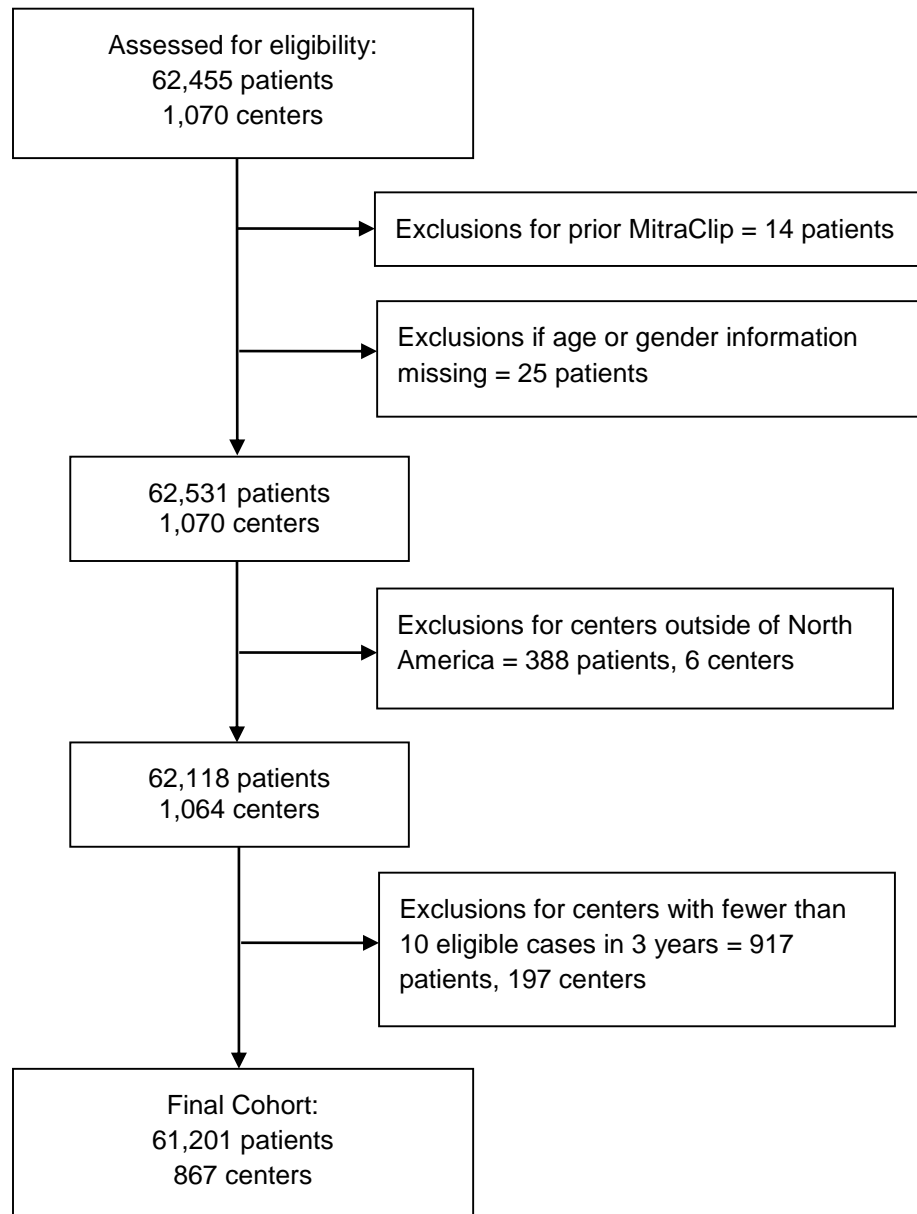


Figure 2

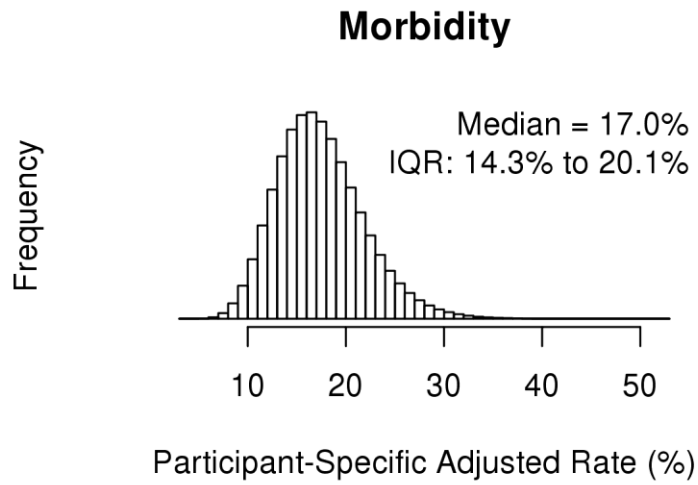
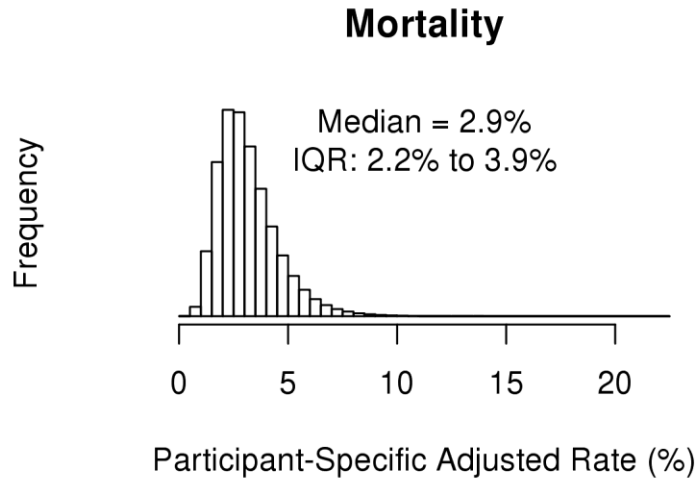


Figure 3

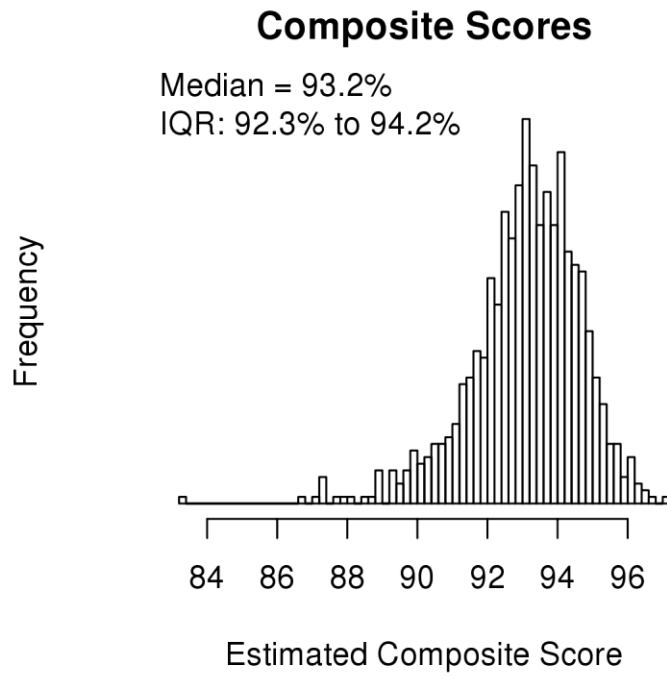
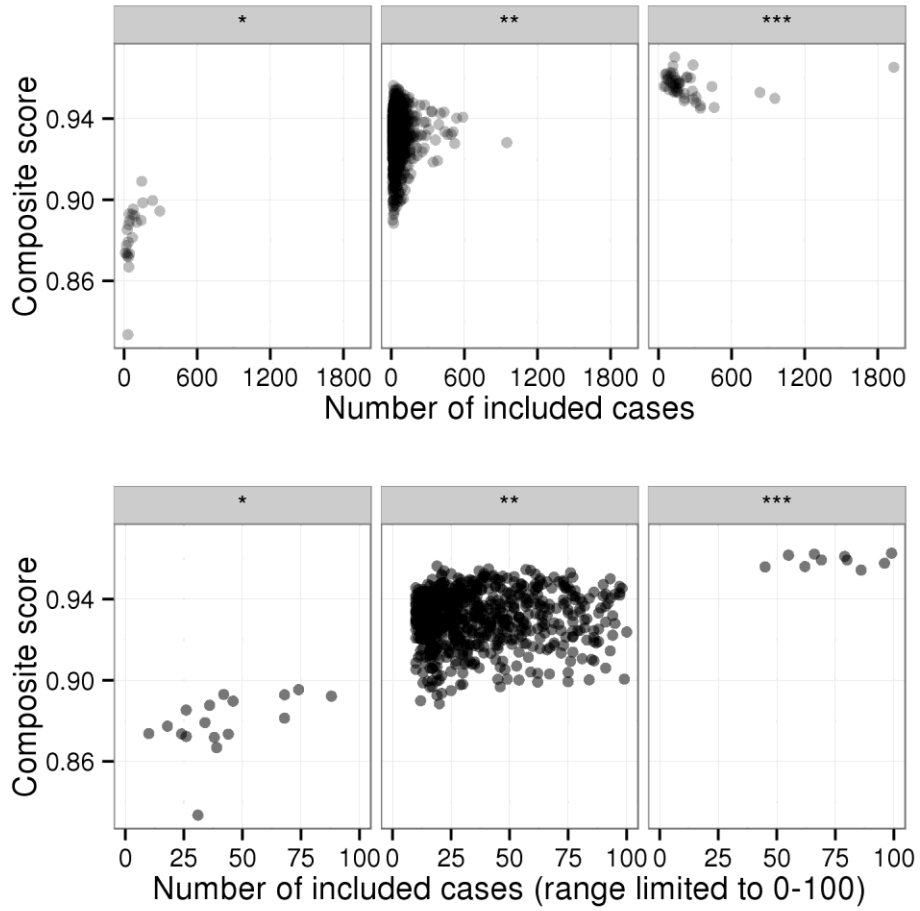


Figure 4



APPENDIX:

A. Statistical Approach to the STS Isolated MVRR Composite

Reliability is conventionally defined as the proportion of variation in a performance measure that is due to true between-hospital differences (i.e., signal) as opposed to random statistical fluctuations (i.e., noise). A mathematically equivalent definition is the squared correlation between a measurement and the true value. This quantity cannot be calculated directly because the “true” composite measure values are unknown, but may be estimated, as described below.

Calculation Details

Let θ_j denote the true unknown composite measure value for the j^{th} of J hospitals. Before estimating reliability, the numeric value of θ_j was estimated for each hospital under the assumed hierarchical model. Estimation was done using Markov Chain Monte Carlo (MCMC) simulations and involved the following steps:

1. For each j , we randomly generated a large number (N) of possible numeric values of θ_j by sampling from the Bayesian posterior probability distribution of θ_j . Let $\theta_j^{(i)}$ denote the i^{th} of these N randomly sampled numeric values for the j^{th} hospital.
2. For each j , a Bayesian estimate $\hat{\theta}_j$ of θ_j was calculated as the arithmetic average of the randomly sampled values $\theta_j^{(1)}, \dots, \theta_j^{(N)}$; in other words $\hat{\theta}_j = 1/N \sum_{i=1}^N \theta_j^{(i)}$.

Our reliability measure was defined as the estimated squared correlation between the set of hospital-specific estimates $\hat{\theta}_1, \dots, \hat{\theta}_J$ and the corresponding unknown true values $\theta_1, \dots, \theta_J$. Let ρ^2 denote the *unknown true* squared correlation of interest and let $\hat{\rho}^2$ denote *an estimate* of this quantity. The estimate was calculated as

$$\hat{\rho}^2 = \frac{1}{N} \sum_{i=1}^N \rho_{(i)}^2$$

where

$$\rho_{(i)}^2 = \frac{[\sum_{j=1}^J (\theta_j^{(i)} - \bar{\theta}^{(i)})(\hat{\theta}_j - \bar{\theta})]^2}{\sum_{j=1}^J (\theta_j^{(i)} - \bar{\theta}^{(i)})^2 \sum_{j=1}^J (\hat{\theta}_j - \bar{\theta})^2} \bar{\theta} = \frac{1}{JN} \sum_{j=1}^J \sum_{i=1}^N \theta_j^{(i)}$$

and

$$\bar{\theta}^{(i)} = \frac{1}{J} \sum_{j=1}^J \theta_j^{(i)}.$$

A 95% Bayesian probability interval for ρ^2 was obtained calculating the 2.5th and 97.5th percentiles of the set of numbers $\rho_{(1)}^2, \dots, \rho_{(N)}^2$.

B. Risk Adjusted Odds Ratio between TVr and no TVr Patients by Degree of Tricuspid Regurgitation at the time of Mitral Valve Repair/Replacement. (TVr = tricuspid valve repair, OR = odds ratio, CI = confidence interval) [11]

Degree of Tricuspid Regurgitation	Morbidity OR (95% CI), p-value	Mortality OR (95% CI), p-value
None-Mild	1.45 (1.27-1.66), p<0.0001	1.15 (0.81-1.63), p=0. 4362
Moderate	1.26 (1.10-1.44), p=0.0008	0.99 (0.79-1.24), p=0. 9357
Severe	1.49 (1.27-1.74), p<0.0001	0.88 (0.66-1.18), p=0. 3817