NATIONAL QUALITY FORUM

National Voluntary Consensus Standards for Patient Outcomes

Measure Summary

Measure number: OT1-013-09

Measure name: The STS CABG Composite Score.

Description: This multidimensional performance measure is comprised of four domains consisting of 11 individual NQF-endorsed cardiac surgery metrics: (1) Operative Care--use of the internal mammary artery; (2) Perioperative Medical Care (use of preoperative beta blockade; discharge beta blockade, antiplatelet agents, and lipid-lowering agents--an "all-or-none" measure); (3) Risk-adjusted Operative Mortality; and (4) Risk-Adjusted Postoperative Morbidity (occurrence of postoperative stroke, renal failure, prolonged ventilation, re-exploration, or deep sternal wound infection--an "any-or-none" measure).

All measures are based on audited clinical data collected in a prospective registry and are risk-adjusted (with the exception of internal mammary artery use and the four perioperative medications). Based on their percentage scores, a 1 (below average), 2 (average), or 3 (above average) star rating is provided for each STS database participant for each performance domain and overall.

Furthermore, the composite score is also deconstructed into its components to facilitate performance improvement activities by providers. This scoring methodology has now been implemented for over two years and has become for many stakeholders the preferred method of evaluating cardiac surgery performance. STS plans to make this report publicly available in the near future. (Additional materials are available upon request)

Numerator statement: Due to the complex methodology used to construct the composite measure, it is impractical to separately discuss the numerator and denominator. The following discussion describes how each domain score is calculated and how these are combined into an overall composite score. Additional documentation is available in the attached article published as a supplement of The Annals of Thoracic Surgery.

Denominator statement: Please see *'Numerator Statement'*

Level of Analysis: Facility/Agency

Type of Measure: Outcome

Data Source: Electronic Health/Medical records, electronic clinical data, registry data, lab data, pharmacy data, paper medical record/flowsheet

Measure developer: The Society for Thoracic Surgeons

Type of Endorsement: (full or time-limited): Full Endorsement (Recommend composite measure without the star reporting system -21, Do not Recommend-0, April 20-21, 2010 Meeting)

NATIONAL QUALITY FORUM

National Voluntary Consensus Standards for Patient Outcomes
Measure Summary

Summary table of TAP ratings of sub criteria and comments:

IMPORTANCE TO MEASURE AND REPORT		
1d Quality construct	High	Quality construct - meaningful composite measure of different aspects of CABG surgery; Conceptual construct - very strong; extensively tested and validated composite measure; Question on evidence for use of discharge beta blockers -- NQF staff will get the evidence justification for the individual measure [1]
1e Conceptual construct	High	
SCIENTIFIC ACCEPTABILTY		
2a specs	High	SCIENTIFIC ACCEPTABILITY: 2a. specifications - complete for composite; some question whether NQF has the most recent specifications for the endorsed measures -- staff will verify; 2b and c - strong reliability and validity testing; 2f - meaningful differences demonstrated- distribution curves for domains and totals; two domains use " all or none" approach; 2i - component justification -- individual correlations low so chose a method that provides useful information to stakeholders; all domains contribute statistical information 2k - does not use differential weighting; OVERALL: a well tested, implemented measure with several publications
2b reliability	High	
2c validity	High	
2d exclusions	High	
2e risk adjustment	High	
2f meaningful differences	High	
2g comparability	High	
2h disparities	Not applicable	
2i component justification	High	
2j component variability	High	

[1] NQF STAFF note: The evidence for the use of beta blockers at discharge after CABG is twofold:
1. Atrial fibrillation is the most common arrhythmia seen following CABG. It occurs in 10% to 65% of patients after cardiac surgery, usually on the second or third postoperative day .It is associated with a 2- to 3-fold increase in postoperative risk for stroke. Patients at risk for postoperative atrial fibrillation have been identified and include those with
COPD, proximal right CAD, prolonged cross-clamp time, atrial ischemia, advanced age, and withdrawal of beta-blockers. ACC/AHA Guidelines for CABG, 2004

a. Postoperative atrial fibrillation was an independent predictor of long-term mortality-at four to five years, survival was approximately 13% worse in patients who developed postoperative atrial fibrillation (OR = 1.5, $p < 0.001$) . Villareal RP, Hariharan R, Liu BC, et al. Postoperative atrial fibrillation and mortality after coronary artery bypass surgery. J Am Coll Cardiol. 2004;43(5):742-748.

b. Atrial fibrillation following CABG is thought to increase length of stay by 1 to 1.5 days. Kim MH, Deeb GM, Morady F, et al. Effect of postoperative atrial fibrillation on length of stay after cardiac surgery (The Postoperative Atrial Fibrillation in Cardiac Surgery study [PACS(2)]. Am J Cardiol 2001;87(7):881-885.

2. Beta blocker use as secondary prevention in patients with significant coronary artery disease. ACC/AHA Guidelines for Secondary Prevention (2006): "Start and continue indefinitely in all patients who have had myocardial infarction, acute coronary syndrome, or left ventricular dysfunction with or without heart failure symptoms, unless contraindicated. I (A)
Consider chronic therapy for all other patients with coronary or other vascular disease or diabetes unless contraindicated. IIa (C)"

2k Differential weighting	High	
2l Missing scores	High	
USEABILITY		
3a distinctive	High	in use; uses clinical data; data provided to providers; may need public education to understand composites in general;
3b harmonization	High	
3c Added value	High	
3d Decomposition	High	
3e State purpose	High	
FEASIBILITY		
4a Data a byproduct of care	High	uses clinical database; audited
4b Electronic	High	
4d Inaccuracies/errors	High	
4e Implementation	High	

Summary of SC ratings of sub criteria and comments:

IMPORTANCE TO MEASURE AND REPORT	
The measure uses data from the STS registry which includes data from over 90% of hospitals performing CABG surgeries and includes 95% of CABG procedure done in the US.	SC Vote on Importance Yes - 22 No - 0
SCIENTIFIC ACCEPTABILITY	
The Committee questioned including the reporting mechanism within the specifications of the measure – atypical for NQF endorsed measures. The measure is calculated to arrive at a point estimate numerical value with confidence intervals as well as the star system. Committee members note the 98% confidence intervals incorporated into the star reporting system – this is higher than the usual 95% confidence intervals used in most measures. The Committee suggests that standardization is important.	SC vote on scientific acceptability Completely - 15 Partially – 6 Minimally – 0 Not at all – 0
USABILITY	
Committee members suggest that various users of the measure	SC vote on usability

should be able to determine how the measure is reported depending on the goals of their program. A Committee member suggested that the one-two-three star system is likely to be misinterpreted by the public as "good-better-best" when one star represents below average performance.	Completely - 0 Partially – 17 Minimally – 3 Not at all – 1
FEASIBILITY	
Provider has to participate in registry; 90 % of CABG are in registry	SC vote on feasibility Completely - 8 Partially – 12 Minimally – 1 Not at all – 0

Summary of Biostatistical review: None

Attachments: None

THE NATIONAL QUALITY FORUM

COMPOSITE MEASURE SUBMISSION FORM
Version 4.0 August 2009

This form will be used by stewards to submit **composite** measures and by reviewers to evaluate the measures.

Measure Stewards: Complete all non-shaded areas of the form. All requested information should be entered directly into this form. The information requested is directly related to NQF's composite measure evaluation criteria and will be used by reviewers to determine if the evaluation criteria have been met. The specific relevant subcriteria language is provided in a Word comment within the form and will appear if your cursor is over the highlighted area.

The measure steward has the opportunity to identify and present the information that demonstrates the measure meets the criteria. Additional materials will only be considered supplemental. Do not rely solely on materials provided at URLs or in attached documents to provide measure specifications or to demonstrate meeting the criteria. If supplemental materials are provided, be sure to indicate specific page numbers/ web page locations for the relevant information (web page links preferred).

For questions about this form, contact the project director at 202-783-1300. Please email this form to the appropriate contact listed in the corresponding call for measures.

Reviewers: Complete all yellow highlighted areas of the form. Evaluate the extent to which each subcriterion is met and then overall, the extent to which each major criterion is met. Provide the rationale for your rating.

Evaluation ratings of the extent to which the criteria are met
H=High (unquestionably demonstrated to meet the criterion)
M=Moderate (demonstrated to moderately meet the criterion)
L=Low (addressed BUT demonstrated to only minimally meet the criterion)
N=No (NOT addressed; OR incorrectly addressed; OR demonstrated to NOT meet the criterion)
NA=Not applicable (only an option for a few subcriteria as indicated)

(for NQF staff use) **NQF Review #:** OT1-013-09 **NQF Project:** Patient Outcome Phase I
Title of Measure: The STS CABG Composite Score
Brief description of measure (*including type of score, measure focus, target population, time, e.g., Percentage of adult patients aged 18-75 years receiving one or more HbA1c tests per year*): This multidimensional performance measure is comprised of four domains consisting of 11 individual NQF-endorsed cardiac surgery metrics: (1) Operative Care--use of the internal mammary artery; (2) Perioperative Medical Care (use of preoperative beta blockade; discharge beta blockade, antiplatelet agents, and lipid-lowering agents--an "all-or-none" measure); (3) Risk-adjusted Operative Mortality; and (4) Risk-Adjusted Postoperative Morbidity (occurrence of postoperative stroke, renal failure, prolonged ventilation, re-exploration, or deep sternal wound infection--an "any-or-none" measure). All measures are based on audited clinical data collected in a prospective registry and are risk-adjusted (with the exception of internal mammary artery use and the four perioperative medications). Based on their percentage scores, a 1 (below average), 2 (average), or 3 (above average) star rating is provided for each STS database participant for each performance domain and overall. Furthermore, the composite score is also deconstructed into its components to facilitate performance improvement activities by providers. This scoring methodology has now been implemented for over two years and has become for many stakeholders the preferred method of evaluating cardiac surgery performance. STS plans to make this report publicly available in the near future. (Additional materials are available upon request)
►**Type of Measure:** ☒ Composite
Select the most relevant priority area(s), quality domain(s), and consumer need(s).

►**National Priority Partners Priority Area** ☐ patient and family engagement ☒ population health ☒ safety ☐ care coordination ☐ palliative and end of life care ☐ overuse

►**IOM Quality Domain** ☒ effectiveness ☒ efficiency ☐ equity ☒ patient-centered ☒ safety ☐ timeliness

►**Consumer Care Need** ☒ Getting Better ☐ Living With Illness ☒ Staying Healthy

CONDITIONS FOR CONSIDERATION BY NQF	
Four conditions must be met before proposed measures may be considered and evaluated for suitability as voluntary consensus standards:	**NQF Staff**
A. The measure is in the public domain or an intellectual property agreement (measure steward agreement) is signed. *Public domain only applies to governmental organizations. All non-government organizations must sign a measure steward agreement even if measures are made publicly and freely available.* ►**Do you attest that the measure steward holds intellectual property rights to the measure and the right to use any aspects of the measure owned by another entity (e.g., component measures, risk model, code set)?** ☒ Yes ►**Measure Steward Agreement** ☒ Signed and Submitted **OR** ☐ Government entity-public domain (If measure steward agreement not signed for non-government entities, do not submit) ►**Please check if either of the following apply:** ☒ Proprietary Measure ☐ Proprietary Complex Measure w/fees	**A** **Y**☒ **N**☐
B. The measure owner/steward verifies there is an identified responsible entity and process to maintain and update the measure on a schedule that is commensurate with the rate of clinical innovation, but at least every 3 years. ☒ Yes (If no, do not submit)	**B** **Y**☒ **N**☐
C. The intended use of the measure includes both public reporting and quality improvement. ►**Purpose:** ☒ Public reporting ☒ Internal quality improvement ☒ Accountability ☒ Accreditation ☒ Payment incentive ☐ Other, describe: (If not intended for both public reporting and quality improvement, do not submit)	**C** **Y**☒ **N**☐
D. The requested measure submission information is complete. Generally, measures should be fully developed and tested so that all the evaluation criteria have been addressed and information needed to evaluate the measure is provided. Measures that have not been tested are only potentially eligible for a time-limited endorsement and in that case, measure owners must verify that testing will be completed within 24 months of endorsement. ►**Testing:** ☒ Fully developed and tested ☐ Testing will be completed within 24 months (If not tested and no plans for testing within 24 months, do not submit) **Component Measures** (*All components of the composite must be either NQF-endorsed or submitted for consideration for NQF endorsement*) ☒ All component measures are NQF-endorsed measures ☐ Some or all component measures are not NQF-endorsed and have been submitted using the online measure submission tool ►**Have NQF-endorsed measures been reviewed to identify if there are similar or related measures?** ☒ Yes (If no, do not submit) *If there are similar or related measures, be sure to address items 3b and 3c with specific information.* ►**Is all requested information entered into this form?** ☒ Yes (If no, do not submit)	**D** **Y**☐ **N**☐
(for NQF staff use) **Have all conditions for consideration been met?** **Staff Notes (if submission returned):** The denominator and numerator should not be combined together under 2a.)	**Met** **Y**☒ **N**☐

Under testing/analysis 2i.) data/sample " The STS composite measure includes all relevant CABG process and outcomes measures currently endorsed by the NQF. We did not conduct statistical analyses to justify the selection of these 11 NQF measures. The justification for them is based on the credibility of the prior NQF process and their evident relationship to the concept of CABG quality. See attached article for details." too vague and needs further clarification by developer. States under usability section 3a) meaningful, understandable, and useful information that the measure is in use but under 3a.) sub-category if used in a public reporting initative or if used in other programs states N/A. Please explain and clarify. Under 2a.) Measure specifications sub-category level of measurement the developer responded "other" and did not detail this further. Needs further clarification. NQF does not have the measure steward agreement.	

1. IMPORTANCE TO MEASURE AND REPORT	
Extent to which the specific measure focus is important to making significant gains in health care quality (safety, timeliness, effectiveness, efficiency, equity, patient-centeredness) and improving health outcomes for a specific high impact aspect of healthcare where there is variation in or overall poor performance. *Measures must be judged to be important to measure and report in order to be evaluated against the remaining criteria.* (composite measure evaluation criteria) If the component measures are determined to meet the importance criteria 1a, 1b, and 1c, then the composite would meet 1a, 1b, and 1c.	**Eval**
(for NQF staff use) **Specific** **NPP goal**: all healthcare organizations and their staff will work collaboratively with patients to reduce preventable emergency department visits	
1d. Purpose/objective of the Composite **►Describe the purpose/objective of the composite measure:** The STS composite CABG measure was designed to fulfill the IOM vision for performance measurement as described in its publication, **Performance Measurement - Accelerating Improvement:** "Composites are a relatively new concept in the area of performance measurement denoting the combination of indicators for several specific measures into a single number… bundling of measures for specific conditions to determine whether all critical aspects of care have been achieved for an individual patient, thereby enhancing measurement to extend beyond tracking performance on separate measures… …this concept represents a turning point and a relatively new challenge for performance measurement…(it) suggests performance goals considerably more stringent than those captured by the usual single variable measures… …a patient-centered approach that takes into account the full constellation of health care needs" In this context, the goal of the STS CABG composite was to create a comprehensive, multidimensional performance score for this most common of all cardiac surgical procedures. It would be easy for most consumers to interpret, but at the same time it would provide granular "drill-down" capability for more methodologically sophisticated stakeholders and for provider PI initiatives. **►Describe the quality construct used in developing the composite:** The quality construct underlying development of the STS CABG composite generally adheres to the classic, multidimensional Donabedian triad of structure, process, and outcomes domains.	**1d** **H**☒ **M**☐ **L**☐ **N**☐
1e. Conceptual construct for quality **►Describe how the component measures are consistent with and representative of the quality construct:** Participation in a systematic database, an NQF-endorsed structural measure of cardiac surgery quality, is implicitly met, as only participants in the STS database will be eligible for a score. The composite includes two process domains: (1) a perioperative medication domain, scored all-or-none, and consisting of 4 NQF-endorsed medications (preoperative beta blockade; discharge beta blockade, antiplatelet agents, antilipid agents); and (2) an operative care process domain, use of at least one internal mammary artery graft (NQF-endorsed). The composite also includes two outcomes domains: (1) risk-adjusted operative mortality (NQF-endorsed); and (2) risk-adjusted morbidity, scored any-or-none (stroke, re-exploration, deep sternal wound infection, renal failure, prolonged ventilation--all NQF-endorsed)	**1e** **H**☒ **M**☐ **L**☐ **N**☐

Staff Notes to Reviewers:	
Reviewer: Was the threshold criterion, Importance to Measure and Report, met? **Rationale:** Quality construct - meaningful composite measure of different aspects of CABG surgery; Conceptual construct - very strong; extensively tested and validated composite measure; Question on evidence for use of discharge beta blockers -- NQF staff will get the evidence justification for the individual measure	**1** **Y**☐ **N**☐
2. SCIENTIFIC ACCEPTABILITY OF MEASURE PROPERTIES	
Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. (composite measure evaluation criteria)	**Eval**
2a. MEASURE SPECIFICATIONS	
In the future, NQF will require measure stewards to provide a URL link to a web page where current detailed specifications can be obtained? ►Do you have a web page where current detailed measure specifications can be obtained? yes ►If yes, provide web page URL: http://www.sts.org/documents/pdf/AdultCVDataSpecifications2.61.pdf 2a. Precisely Specified **Components of the Composite** (*List the components, i.e., domains/sub-composites and individual measures*) ►**List components:** (*If component measures NQF-endorsed, include NQF measure number*; if not NQF-endorsed, provide date of submission to NQF) The composite consists of four domains and 11 individual measures, all NQF-endorsed: a. Perioperative medication domain, scored all-or-none and consisting of: NQF # 0127 - Pre-Operative Beta Blockade NQF # 0117 - Beta Blockade at Discharge NQF # 0116 - Anti-Platelet Medication at Discharge NQF # 0118 - Anti-Lipid Treatment Discharge b. Operative care process domain NQF # 0134 - Coronary artery bypass graft (CABG) using internal mammary artery (IMA) c. Risk-adjusted operative mortality NQF # 0119 - Risk-Adjusted Operative Mortality for CABG d. Risk-adjusted morbidity, scored any-or-none and consisting of: NQF # 0131 - Stroke/Cerebrovascular Accident NQF # 0115 - Surgical Re-exploration NQF # 0130 - Deep Sternal Wound Infection Rate NQF # 0114 - Post-operative Renal Failure NQF # 0129 - Prolonged Intubation (ventilation)	**2a-specs** **H**☒ **M**☐ **L**☐ **N**☐
Composite Numerator Statement: Due to the complex methodology used to construct the composite measure, it is impractical to separately discuss the numerator and denominator. The following discussion describes how each domain score is calculated and how these are combined into an overall composite score. Additional documentation is available in the attached article published as a supplement of The Annals of Thoracic Surgery. **Numerator Time Window:** The STS composite score currently is based on one year of data. However, we would request that NQF endorsement not be limited to this time window as alternative sampling periods may be employed in the future. **Numerator Details:** Technical Details: The unit of measurement for the STS Composite Score can be either a participant (most often a cardiac surgical practice but occasionally an individual surgeon) or a hospital. The STS composite score is an aggregate of 4 scores corresponding to 4 domains of CABG quality (mortality,	

morbidity, operative care, perioperative medical care). Each domain score has a theoretical range of 0 to 1 and is interpreted as a probability. A description of these probabilities is presented in Table 1 below. Larger values imply better performance. Although the theoretical range of each score (probability) is 0 to 1, the actual scores tend to be clustered in the upper end of the 0-1 interval. For reporting purposes, the probabilities are expressed as percentages ranging from 0% to 100%.

Table 1

#	Domain	Interpretation of Domain Score
1	Risk-adjusted operative mortality	π_1 = The probability (risk-adjusted) that a patient will be discharged alive and will survive to >30 days post-surgery.
2	Postoperative risk-adjusted major morbidity	π_2 =The probability (risk-adjusted) that a patient will be discharged without experiencing any of the following endpoints: stroke/cerebrovascular accident, surgical re-exploration, deep sternal wound infection, post-operative renal failure, prolonged intubation (ventilation).
3	Operative care	π_3 =The probability that a patient without a prior CABG will receive an IMA.
4	Perioperative medical care	π_4 =The probability that a patients will receive all of the medications for which the patient is eligible from the following list: preoperative beta blockade; discharge beta blockade, antiplatelet agents, antilipid agents.

Separate probability estimates are calculated for each unit in the analysis. The method of Bayesian multivariate hierarchical regression modeling is used to obtain estimates that account for chance variation and noisy data. The Bayesian statistical framework is used to assess statistical significance, calculate measures of uncertainty, and distinguish true variation from random noise. After estimating the probability parameters for a unit, a composite score is calculated for each unit by using the following formula:

$$\text{STS composite score} = \frac{\hat{\pi}_1}{\text{sd}_1} + \frac{\hat{\pi}_2}{\text{sd}_2} + \frac{\hat{\pi}_3}{\text{sd}_3} + \frac{\hat{\pi}_4}{\text{sd}_4}$$

where $\hat{\pi}_1$, $\hat{\pi}_2$, $\hat{\pi}_3$, $\hat{\pi}_4$ denote the estimate (for the unit of interest) of the probabilities defined in Table 1 above and sd_1, sd_2, sd_3, sd_4 denote the standard deviations of these estimate across all units in the analysis. In the most recent production of the STS composite measure based on data from 2008, the estimated standard deviations were approximately $\text{sd}_1 = 0.523$, $\text{sd}_2 = 5.90$, $\text{sd}_3 = 4.98$, $\text{sd}_4 = 15.4$.

Estimation of the STS composite score is based on unit-level summary data rather than patient-level data. The required data consist of numerators and denominators for each of the four domain scores, as described in Table 2.

Table 2

Domain	Denominator	Numerator
Mortality	The number of patients undergoing isolated coronary artery bypass grafting (CABG) during the measurement period.	The number of these patients who survived until after discharge and >30 days post-surgery.
Morbidity	The number of patients undergoing isolated coronary artery bypass grafting (CABG) during the measurement period.	The number of these patients who did not experience any of the selected morbidity endpoints[1].
Operative care	The number of patients undergoing isolated coronary artery bypass grafting (CABG) during the measurement period who did not have a prior CABG.	The number of these patients who received an IMA.
Medications	The number of patients undergoing isolated coronary artery bypass grafting (CABG) during the measurement period who were eligible to receive at least one NQF perioperative medication[2].	The number of these patients who received all of the medications for which the patient was eligible[2].

1. Morbidity endpoints consist of stroke/cerebrovascular accident, surgical re-exploration, deep sternal wound infection, post-operative renal failure, prolonged intubation (ventilation). Patients with prior CVA are excluded when counting stroke outcomes. Patients with history of renal failure (creatinine > 2.0) are excluded when counting renal failure outcomes.

2. Medications consist of: preoperative beta blockade; discharge beta blockade, antiplatelet agents, antilipid

agents. For preoperative betablockers, all patients without a documented contraindication are considered to be eligible. For discharge medications, patients without a documented contraindication who are alive at the time of discharge are considered to be eligible.

In addition to the numerators and denominators defined in Table 2, two additional items are required for calculating the composite measure. These are:

1. The average predicted risk of operative mortality for patients undergoing isolated CABG during the measurement period, as determined by the current STS CABG mortality model.

2. The average predicted risk of "mortality or major morbidity" for patients undergoing isolated CABG during the measurement period, as determined by the current STS CABG mortality/major morbidity model.

The current version of the STS CABG risk models can be found in the following article:

Shahian DM, O'Brien SM, Filardo G, et al. The Society of Thoracic Surgeons 2008 Cardiac Surgery Risk Models: Part 1—Coronary Artery Bypass Grafting Surgery. Ann Thorac Surg 2009;88:S2-S22.

Composite Denominator Statement: Please see response in numerator statement above

Denominator Time Window:

Denominator Details:

Composite Denominator Exclusions: Please see response in numerator statement above

Denominator Exclusion Details:

►**Type of Score:** Non-weighted score/composite/scale ► If "Other", please describe:

►**Interpretation of Score** (*Classifies interpretation of score according to whether better quality is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score*)
Better quality = Higher score ► If "Other", please describe:

Method of Scoring/Aggregation: other **If "other" scoring method, describe:** The operative mortality score is a risk-adjusted rate. The morbidity domain is a risk-adjusted rate, and requires absence of all 5 major complications. IMA use is an unadjusted proportion. Perioperative medication use is scored all-or-none. Please refer to composite numerator statement for overall aggregation methodology.

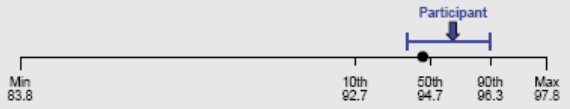
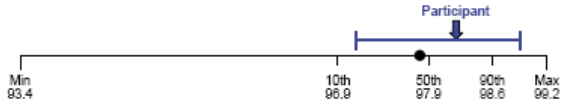
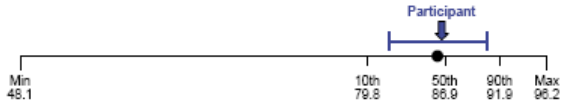
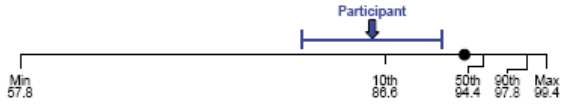
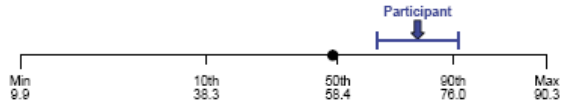
Missing Component Scores (*Indicate how missing component scores are handled*): An overall composite score is not calculated if any domain measures are missing

Weighting: ☒ **Equal** ☐ **Differential** **If differential weighting, describe:**

►**Calculation Algorithm** (*Describe the calculation of the measure as a flowchart or series of steps*)**:**
Please see discussion under composite numerator statement and attached articles.

►**Describe the method for discriminating performance** *(e.g., significance testing)***:**
STS Composite Quality Rating Interpretation Example

Page 1: Scores and Star Ratings Sample Results

Quality Domain	Participant Score (98% CI)	STS Mean Participant Score	Participant Rating[1]	Distribution of Participant Scores • = STS Mean
2006 Overall	95.3% (94.1 , 96.3)	94.5%	★★	
2006 Avoidance of Mortality	98.2% (97.1 , 98.9)	97.8%	★★	
2006 Avoidance of Morbidity[2]	86.6% (81.8 , 90.7)	86.2%	★★	
2006 Use of IMA[3]	85.6% (80.0 , 91.1)	92.9%	★	
2006 Medications[4]	70.6% (64.3 , 76.7)	57.6%	★★★	

Explanation of Report Page 1:

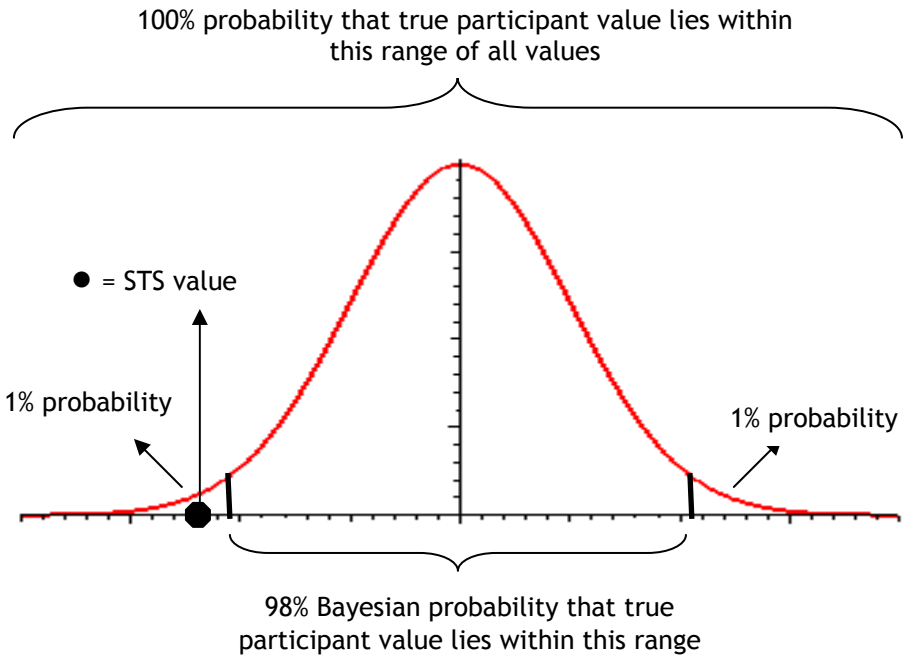
Column 1. Quality Domain**.** The quality domain for which results are provided. See table above for more information about the four quality domains.

Column 2. Participant Score and 98% Credible Interval.

Participant Score. The participant score is a number that summarizes the participant's estimated performance within the indicated domain of quality. Scores are expressed as percentages with higher numbers indicating better quality. The score may be interpreted as the participant's estimated underlying success rate for the endpoint that was used to measure performance within the domain. The underlying success rate cannot be observed directly, but may be estimated from the data. It is interpreted as the success rate that would be observed hypothetically if the participant treated a very large number of patients (so that the observed results were not subject to sampling variation.) For example, a score of 85.6% for the IMA domain indicates that the estimated underlying success rate for IMA usage is 85.6%. This means that approximately 85.6% of first-time CABG patients would be projected to receive an IMA in the long run if the participant treated a very large population. Success rates for the mortality and morbidity domain are risk-adjusted. The participant score is interpreted as the proportion of patients who would not experience the indicated adverse outcome if the participant treated a large number of patients having a

"typical" case mix. The calculation and interpretation of scores within each domain are described in greater detail above.

98% Bayesian Credible Intervals. The 98% Bayesian credible interval shows the range in which the participant's true proportion (or risk-adjusted proportion) is likely to lie. Unlike traditional confidence intervals, the Bayesian credible interval has an intuitive probability interpretation. For example, based on the observed data, we can state that it is 98% likely that the true proportion (or risk-adjusted proportion) falls within the upper and lower limits of the 98% credible interval. If the lower limit of the 98% Bayesian credible interval is greater than the STS average value, then it is at least 99% likely (98% credible interval plus the 1% upper tail) that the participant's true performance exceeds the STS average value. The figure below illustrates this. Conversely, if the upper limit of the 98% Bayesian credible interval is less than the STS value, then it is at least 99% likely (98% credible interval plus 1% lower tail) that the participant's true performance is less than (i.e., worse than) the STS value.



Column 3. STS Mean Participant Score. The STS mean participant score is the average of all scores across all of the participants in the analysis. This score serves as a useful benchmark for assessing a participant's performance relative to the overall STS performance.

Column 4. Participant Rating. The participant rating system assigns participants to rating categories designated by one, two, or three stars. The rating categories are defined as follows:

★★★ → Participant performance is significantly higher than STS mean.

★★ → Participant performance is not statistically different from STS mean.

★ → Participant performance is significantly lower than STS mean.

Statistical significance is based on a 99% Bayesian certainty criterion. A participant receives 3 stars if there is at least 99% Bayesian probability that the participant's score exceeds the STS mean score. A participant receives 1 star if there is at least 99% Bayesian probability that the participant's score is less than the STS mean score. Otherwise, the participant receives 2 stars.

Column 5. Distribution of Participant Scores. This graphical display summarizes the distribution of participant scores across all participants who were included in the analysis. The labels "Min" and "Max" denote the estimated lowest and highest participant scores, among all participants in the analysis. The labels "10th", "50th", and "90th" denote the 10th, 50th, and 90th percentiles of the distribution across participants. (The p-th percentile is a number having the property that p% of participants have scores that are lower than this number and [100-p]% of participants have scores that exceed this number.) The STS mean participant score is indicated by a dot. The participant's score is indicated by an arrow and is accompanied by a horizontal bar representing the limits of the 98% Bayesian credible interval.

Page 2: Quality Domain Details Sample Results

Quality Domain	Eligible Procedures	Detail	Count	Percent of Morbidity/Failure[1]
2006 Avoidance of Mortality	553	Mortality	6	
2006 Avoidance of Morbidity[2]	553	Any Morbidity	73	
		Reoperation only[3]	14	19.2%
		Renal Failure only[4]	8	11.0%
		Deep Sternal Wound Infection only	0	0.0 %
		Prolonged Ventilation only	33	45.2%
		Cerebrovascular Accident only[5]	1	1.4 %
		Multiple Morbidities	17	23.3%
2006 Use of IMA[6]	522	IMA Failures	75	
2006 Medications[7]	553	Failed to Prescribe all eligible NQF Medications	158	
		Only failed to prescribe Preoperative Beta Blockade	70	44.3%
		Only failed to prescribe Discharge Beta Blockade[8]	17	10.8%
		Only failed to prescribe Discharge Anti-Lipids[8]	31	19.6%
		Only failed to prescribe Discharge Anti-Platelets[9]	16	10.1%
		Failed to prescribe multiple medications	24	15.2%

Explanation of Report Page 2:

Column 1. Quality Domain. The quality domain for which results are provided. See table above for more information about the four quality domains.

Column 2. Eligible Procedures. Denotes the number of procedures that were included in the denominator when estimating the success rate for the avoidance of mortality and morbidity endpoints and the adherence to process measures (medications and IMA use).

Column 3. Detail. For those measures that comprise a composite of multiple NQF measures, the specific measures are listed individually to provide detail for quality improvement initiatives.

Column 4. Count. The number of records for which the mortality, morbidity, or process failure occurred. For the composite morbidity and medication domains, counts are provided of the number of records having each of the component failures, as well as the number of records having multiple component failures.

Column 5. Percent of Morbidity/Failure. The proportion that the specific morbidity or process non-compliance contributed to the total number of patients for whom credit was not received for these 'all/any or none' bundles. This information is intended to facilitate and focus process and quality improvement initiatives by providers.

►**Sampling (Survey) Methodology** *If measure is based on a sample (or survey), provide instructions for obtaining the sample, conducting the survey and guidance on minimum sample size (response rate)*:
This is not a sample. All CABG patients cared for by a provider in a given time period are used to calculate the composite score.

►**Stratification Details/Variables (***All information required to stratify the measure including the stratification variables, all codes, logic, and definitions***):**
Not stratified

►**Data Source** *Check all the source(s) used in the component measures.*

- ☐ Electronic administrative data/ claims
- ☒ Electronic Health/Medical Record
- ☒ Electronic Clinical Data *(e.g., MDS)*
- ☒ Registry data (or database)
- ☒ Lab data
- ☒ Pharmacy data
- ☒ Paper Medical Record/flowsheet
- ☐ Survey-patient *(e.g., CAHPS)*
- ☐ Survey-provider
- ☐ Documentation of original self-assessment *(e.g., SF-36)*
- ☐ Management data
- ☐ Public health data/vital statistics
- ☐ Special or unique data, specify:

►**Level of Measurement/Analysis** (*For what entity will the scores be computed?*)
Check the level(s) for which the measure is specified and tested.

Clinician: ☐ Individual ☒ Group ☐ Other
- ☒ Facility/Agency (e.g., hospital, nursing home)
- ☐ Multi-site/corporate chain
- ☐ Integrated delivery system
- ☐ Health plan
- ☐ Prescription drug plan

Program: ☐ Disease management ☐ QIO
- ☒ Other

Population: ☐ National ☐ Regional/network
- ☐ State ☐ Counties/Cities
- ☐ Other (*Please describe*):
- ☒ All levels

►**Applicable Care Settings**
Check the setting(s) for which the measure is specified and tested.
Ambulatory Care: ☐ Amb Surgery Center ☐ Office ☐ Clinic ☐ Emergency Dept ☐ Hospital Outpatient

- ☐ Assisted Living
- ☐ Behavioral health/psychiatric unit
- ☐ Dialysis Facility
- ☐ Emergency medical services/ambulance
- ☐ Group Home
- ☐ Home
- ☐ Hospice
- ☒ Hospital
- ☐ Long term acute care hospital
- ☐ Nursing home/ Skilled Nursing Facility (SNF)
- ☐ Rehabilitation Facility
- ☐ Other (*Please describe*):
- ☐ Unspecified or "not applicable"
- ☐ All settings

TESTING/ANALYSIS

2i. Component item/measure analysis to justify inclusion in composite

Data/sample: The STS composite measure includes all relevant CABG process and outcomes measures currently endorsed by the NQF. We did not conduct statistical analyses to justify the selection of these 11 NQF measures. The justification for them is based on the credibility of the prior NQF process and their evident relationship to the concept of CABG quality. See attached article for details. Wherever possible, we applied classical *psychometric* principles to the development of the composite domains. However, as the four domains may measure multiple distinct aspects of CABG quality, the overall measure also may be viewed from a *clinimetric* perspective.

Data/sample: Empirical analyses were performed using data from the STS database for isolated CABG surgeries between January 1 – December 31, 2004. The analysis was restricted to 530 database participants that performed at least 10 isolated CABG surgeries during 2004 and had less than 5% missing data for each of the five NQF process measures.

Analytic Method: The correlation between individual NQF items was assessed empirically but not for the purpose of including or excluding items. Our purpose was to shed light on the appropriate analytic strategy. If several items are highly correlated, an appealing option for combining them into a single composite measure is to perform a latent trait analysis. This method of analysis relies on a key assumption, namely that several individual items are all related to a single underlying latent variable (i.e. the assumption of unidimensionality). High inter-item correlations suggest that the assumption of unidimensionality may be plausible, whereas low correlations suggest that alternative models may be more appropriate.

Testing Results: In a pilot study using STS CABG data from 2004, we found that the individual NQF process and

2i
H☒
M☐
L☐
N☐

outcome measures were only moderately correlated. Pearson correlation coefficients ranged from 0.10 to 0.50 for pairs of process measures and from 0.15 to 0.61 for pairs of outcome measures. In a related analysis, we found that the assumptions underlying a latent trait logistic regression model were not satisfied for the STS data set. These empirical analyses and our interpretation of them are described in Part 2 of our attached report (See Tables 2 and 3; text on page 16, column 2; and text on page 17, part d).

Interpretation: These results led us to select a composite measure methodology that did not rely on the assumption of unidimensionality. The finding that individual NQF measures were not highly correlated led to important caveats: (1) the choice of how to weight the individual items would be difficult to justify scientifically; (2) different weights may lead to different results; (3) the practical implications of different weighting schemes may be difficult to anticipate. In light of these limitations, our goal was to create a composite measure that would be regarded as useful and valid by key stakeholders. The implications and behavior of our chosen methodology were explored both analytically and empirically and limitations were reported with full transparency. As described below, response from users of the composite measure has been highly positive with no major issues or shortcomings identified.

2j. Component item/measure analysis of contribution to variability in composite score

Data/sample: Isolated CABG surgeries captured in the STS database between January 1 – December 31, 2004. The analysis was restricted to 530 database participants that performed at least 10 isolated CABG surgeries during 2004 and had less than 5% missing data for each of the five NQF process measures.

Analytic Method: The STS composite measure is a combination of 4 separate domain-specific scores. To verify that each domain contributes statistical information but does not dominate the composite, we calculated the [item-total] correlation between each domain-specific estimate and the overall comprehensive score.

Testing Results: The [item-total] Pearson correlations were 0.48 (IMA score versus overall score), 0.56 (medication domain score versus overall score), 0.65 (morbidity domain score versus overall score), and 0.78 (mortality domain score versus overall score).

Interpretation: Although risk-adjusted mortality and morbidity explain much of the variation in the overall comprehensive score, no single item dominates, and all four items contribute statistical information.

2j
H☒
M☐
L☐
N☐

2k. Analysis to support differential weighting of component scores

Data/sample: The composite measure did not use differential weighting. The overall composite score is an equally weighted average of scores from 4 domains of quality. The domain scores are standardized to have a common standard deviation before averaging them equally. This standardization was used to ensure that items with large standard deviations (medications) would not dominate the overall composite score. Results were nearly identical when standardization was accomplished by dividing by the range instead of the standard deviation. Please see Column C in the table below.

	A Perioperative Medical Care Domain All or None Versus CMS or Simple Average	B Postoperative Morbidity Domain Any or None Versus Simple Average	C Overall Composite Score Rescale by SD Versus Range	D Overall Composite Score Rescale by SD Versus No Rescaling
Spearman rank correlation	0.98	0.95	0.99	0.84
Rank changes by				
>50 places	8.7%	24.9%	2.1%	50.6%
>100 places	0.8%	4.5%	0.0%	24.5%
>200 places	0.0%	0.0%	0.0%	3.2%
Top 1/3 by one, not other	10.2%	11.9%	5.7%	25.0%
Top 1/3 by one, bottom 1/3 by other	0.0%	0.0%	0.0%	2.3%

CMS = Centers for Medicare & Medicaid Services; SD = standard deviation.

2k
H☒
M☐
L☐
N☐

Analytic Method:

Testing Results:

Describe how the method of scoring/aggregation achieves the stated purpose and represents the quality construct:

Indicate if any alternative scoring/aggregation methods were tested and why not chosen:

2l. Analysis of missing component scores

Data/sample: In the current STS implementation, composite scores are only calculated for participants with at least 95% complete data for NQF process measures. Missing data regarding care processes for individual patient records are imputed. Patients with missing data are assumed not to have received the care process. This imputation strategy is intended to create an incentive for participants to collect complete process measure data. Sensitivity to the choice of missing data method was assessed at the time of measure development using 2004 STS data. Inferences about provider performance were virtually identical when missing data were imputed to the worst care value vs. excluded from the analysis.

For each of the five NQF process measures, we calculated each hospital's measure two ways (first imputing missing data, then excluding missing data) and compared the results. In each case, the Pearson correlation between the two versions of the process measure was >0.995.

Analytic Method:

Testing Results:

2l
H☒
M☐
L☐
N☐

2b. Reliability testing of composite score

►**Data/sample** *(description of data/sample and size)*: Pilot testing was performed using data from the STS database for isolated CABG surgeries between January 1 – December 31, 2004. The analysis was restricted to 530 database participants that performed at least 10 isolated CABG surgeries during 2004 and had less than 5% missing data for each of the five NQF process measures.

►**Analytic Method** *(type of reliability & rationale, method for testing)*: We assessed whether a single year of data was sufficient to detect statistically significant performance differences between STS participants based on the composite measure. We determined the number and proportion of STS participants that were classified as "outliers" (in the sense that their performance was statistically distinguishable from the overall STS average). For comparison, we determined the number and proportion of STS participants that were classified as "outliers" when outlier status was based on risk-adjusted operative mortality instead of the composite measure. In each case, outlier status was based on a Bayesian posterior probability criterion. Participants were labeled as "better than average outliers" if it was at least 99% certain that the participant's true composite score (or mortality rate) was better than the overall STS average composite score (or mortality rate). Participants were labeled as "worse than average outliers" if it was at least 99% certain the participant's true composite score (or mortality rate) was worse than the overall STS average composite score (or mortality rate).

►**Testing Results** *(reliability statistics, assessment of adequacy in the context of norms for the test conducted)*: The number of participants classified as "better than average" or "worse than average" using this particular rating system with 2004 data was 70 (13%) and 53 (10%), respectively. For these 123 providers, the classification of above average or below average performance could be made with high confidence (more than 99% Bayesian certainty). In comparison, only 6 participants (1%) could be identified as better or worse than average using the same rigorous criterion of at least 99% certainty.

Interpretation: These results demonstrate that the composite measure achieves high power relative to the more commonly used metric of risk-adjusted mortality. There is substantial true signal variation in the endpoints that contribute to the composite measure. Moreover, one year of data is sufficient to identify several outliers. Naturally, the number of performance outliers would be even larger if we used a less rigorous criterion for assessing

2b
H☒
M☐
L☐
N☐

statistical significance.

2c. Validity testing of composite score

►**Data/sample** *(description of data/sample and size)*: 2004 pilot study data and five subsequent semiannual reports distributed to STS participants and stakeholders.

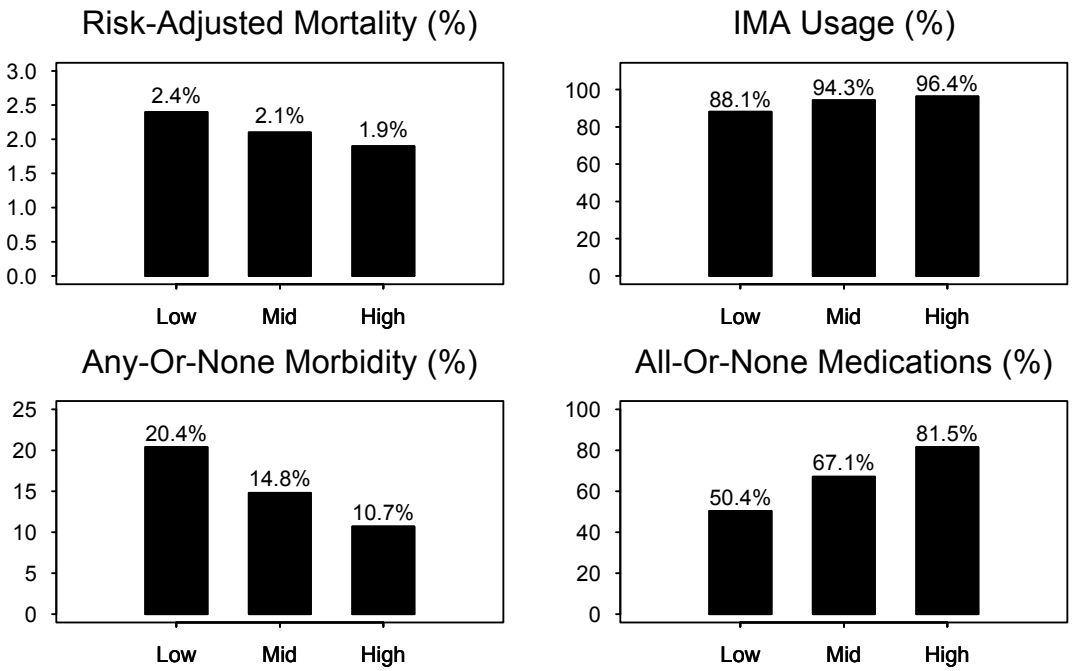
►**Analytic Method** *(type of validity & rationale, method for testing)*: The validity of the composite measure was assessed from multiple perspectives including construct validity, content validity, predictive validity, attributional validity, and face validity. Construct validity implies that the composite measure is a faithful reflection of the underlying concept of CABG quality. As previously discussed, our STS CABG composite is based on the Donabedian quality construct of structure, process, and outcomes. All three of these areas are encompassed by the STS composite (structure is implicitly represented, as only programs participating in our systematic data registry will receive a score). Content validity means that the composite measure includes all of the essential dimensions of the underlying concept. STS believes that the eleven individual NQF-endorsed measures in the composite are both inclusive and broadly representative of the latent construct "CABG quality". Predictive validity means that the results of this measure are predictive of future performance. This was assessed empirically as described below in "temporal stability and predictive validity." Attributional validity means that adequate risk-adjustment has been employed so that differences in quality are not biased by differences in patient-severity. Based on the extensive series of articles we have provided describing the development and testing of STS risk models, we believe this criterion has been satisfied. Face validity implies that the measure is regarded as useful and valid by its intended users, including providers, consumers, payers, and regulators. Now in our third year of actual implementation, we have had near-universal acceptance of this composite by all stakeholders, with few if any relevant suggestions for change.

►**Testing Results** *(statistical results, assessment of adequacy in the context of norms for the test conducted)*: Participants were labeled as "better than average outliers" if it was at least 99% certain that the participant's true composite score (or mortality rate) was better than the overall STS average composite score (or mortality rate). Participants were labeled as "worse than average outliers" if it was at least 99% certain the participant's true composite score (or mortality rate) was worse than the overall STS average composite score (or mortality rate). Participants were labeled as better than average (3 stars), worse than average (1 star), and indistinguishable from the average (2 stars).

A. Construct and Content Validity

STS participants with high composite scores have (on average) higher performance on each individual domain of the composite measure. Thus, differences in performance were clinically meaningful as well as statistically significant. Compared to participants receiving 1 star, those with 3 stars had better estimated performance for each individual domain of the composite score. This is illustrated in the figure below using data from Fall 2008. Participants with 1, 2, and 3 stars are denoted by the labels "low", "mid", and "high" respectively. Compared to participants receiving 1 star, those with 3 stars had lower risk-adjusted mortality (1.9% vs. 2.4%), lower morbidity (10.7% vs. 20.4%), higher IMA usage (88.1% vs. 96.4%), and higher all-or-none medication adherence (81.5% vs. 50.4%).

2c
H☒
M☐
L☐
N☐



B. Predictive Validity

We assessed the extent to which performance on the STS composite measure remains stable over time. In other words, does the composite score performed at one point in time accurately predict performance at some later time? The analysis was restricted to a sample of 706 STS participants who consistently participated and received a composite score in each of the 4 harvest periods (Spring 2007, Fall 2007, Spring 2008, Fall 2008). Among participants who received a 3-star rating in Spring 2007, 51% of them also received a 3-star rating for Fall 2008. For comparison, only 9% of participants who received a 2-star rating in Spring 2007 received a 3-star rating in Fall 2008. Thus, participants who performed above average in Spring 2007 were over 5 times more likely to be identified as above average 1.5 years later. Similarly, participants who were 1 star in Spring 2007 were more likely to be 1 star in 2008. Only 2 participants changed from 1 star to 3 star status (or vice versa) between the two time period. Overall, change by one star class occurred about half the time, but change of two star classes was extremely rare. Thus, a consumer may reasonably expect that an above or below average score will likely be the same or average in the near future, and an average score is likely to remain average about 85% of the time.

Table 3. Change in star ratings between 1st and 4th harvests periods.

Spring 2007	Fall 2008		
	1	2	3
	N	N	N
1	44	44	1
2	31	434	46
3	1	51	54

2f. Identification of Meaningful Differences in Performance

►**Data/sample from Testing or Current Use** *(description of data/sample and size)*: Since 2007, the STS composite score has been reported in bi-annual feedback reports to STS database participants. We summarized star rating results for the first 4 bi-annual STS reports.

►**Methods to identify statistically significant and practically/meaningfully differences in performance** *(type of analysis & rationale)*: In the current STS implementation, the degree of uncertainty surrounding an STS participant's composite measure estimate is indicated by calculating Bayesian credible intervals (CI's) which are

2f
H☒
M☐
L☐
N☐

similar to conventional confidence intervals. Point estimates and CI's for an individual STS participant are reported along with a comparison to various benchmarks based on the national sample. Benchmarks include the overall average STS composite score and several percentiles(minimum, 10th, 25th, 75th, 90th, maximum). A sample of the current STS reporting format is provided in the Appendix. In addition, the composite measure result is converted into a star rating of 1 to 3 stars. An STS participant receives 2 stars if the Bayesian credible interval surrounding their composite score overlaps the overall STS average. This rating implies that the STS participant's performance was not statistically different from the overall STS national average. If the Bayesian CI falls entirely above the STS national average, the participant receives 3 stars (better than average performance). If the Bayesian CI falls entirely below the STS national average, the participant receives 3 stars (worse than average performance).

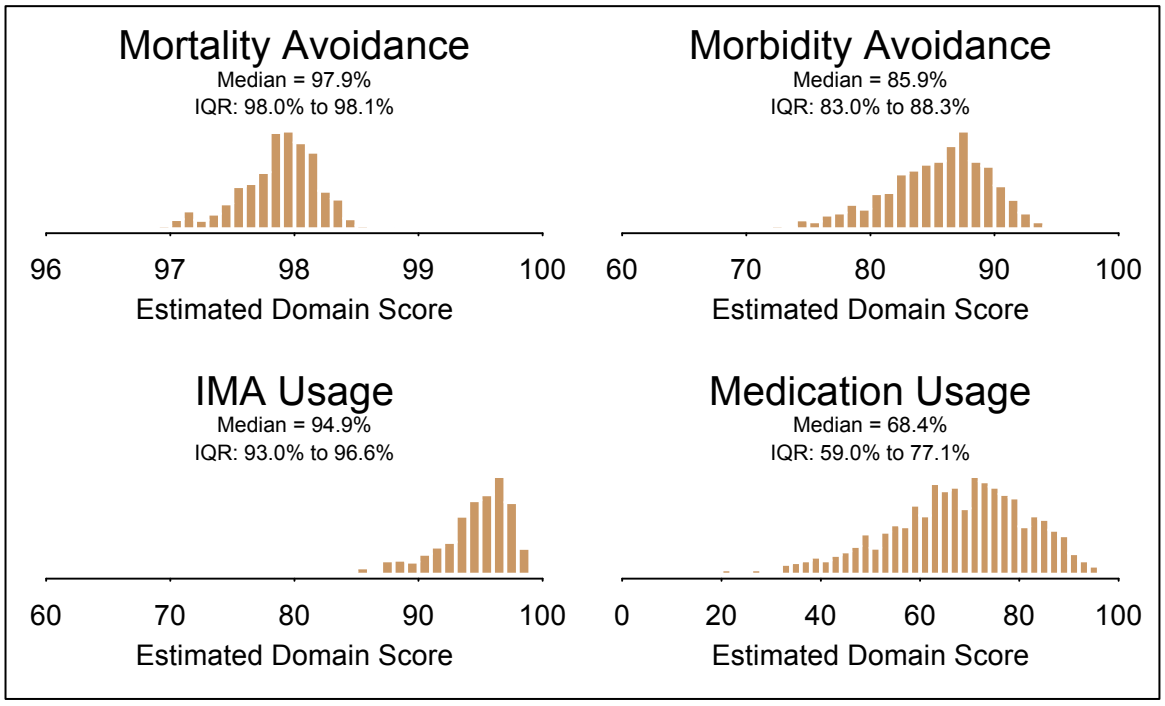
As shown in the Table 4 below, the proportion of STS participants receiving 1, 2, or 3 stars has remained roughly constant over the first 4 reporting periods. On average, roughly 3/4 of participants have received 2 stars, and the remaining 1/4 of participants have received either 1 or 3 stars.

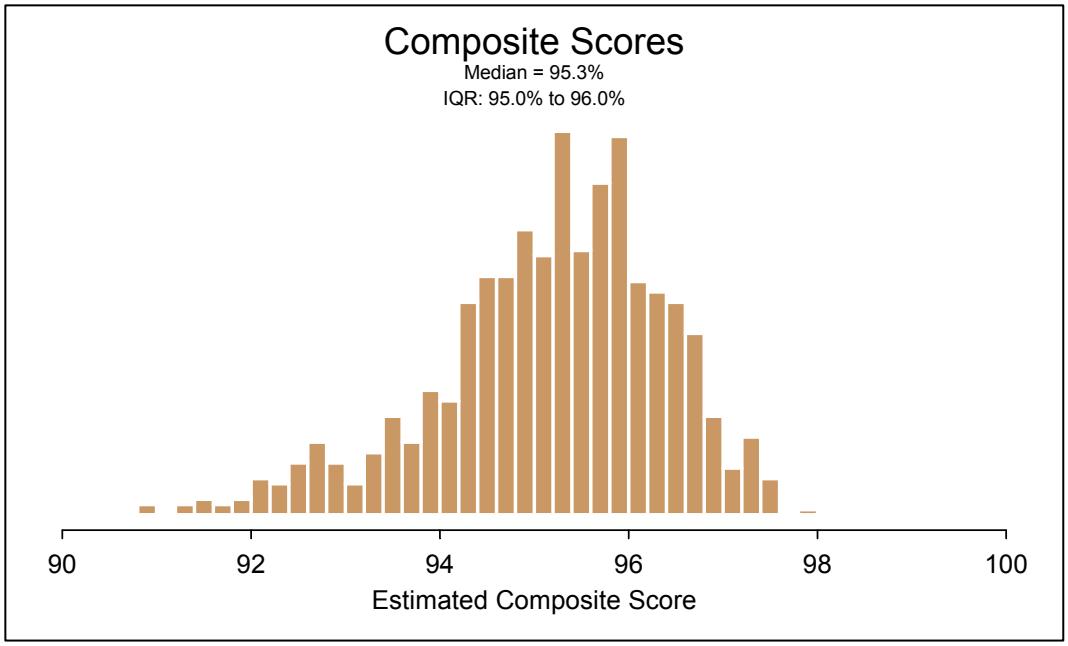
Table 4. Proportion of participants in each star rating category by harvest.

Star Rating	Spring 2007	Fall 2007	Spring 2008	Fall 2008	Spring 2009
1	12.9%	11.4%	12.1%	12.1%	12.2%
2	72.0%	76.5%	76.2%	74.1%	72.3%
3	15.1%	12.1%	11.6%	13.9%	15.5%

Average number of participants for each harvest was 818 and included data on an average of approximately 154,602 patients.

► **Provide Measure Scores from Testing or Current Use** *(description of scores, e.g., distribution by quartile, mean, median, SD, etc.; identification of statistically significant and meaningfully differences in performance)* **:**





2h. Disparities in Care

►**If measure is stratified, provide stratified results** *(scores by stratified categories/cohorts)*:
Not stratified. N/A

►**If disparities have been reported/identified, but measure is not specified to detect disparities, provide follow-up plans:**

2h
H☐ M☐ L☐ N☐ NA☒

Staff Notes to Reviewers:

Reviewers: Overall, to what extent was the criterion, *Scientific Acceptability of Measure Properties*, **met?**
Rationale: SCIENTIFIC ACCEPTABILITY: 2a. specifications - complete for composite; some question whether NQF has the most recent specifications for the endorsed measures -- staff will verify; 2b and c - strong reliability and validity testing; 2f - meaningful differences demonstrated- distribution curves for domains and totals; two domains use " all or none" approach; 2i - component justification -- individual correlations low so chose a method that provides useful information to stakeholders; all domains contribute statistical information 2k - does not use differential weighting;
OVERALL: a well tested, implemented measure with several publications

2
H☒ M☐ L☐ N☐

3. USABILITY

Extent to which intended audiences (e.g., consumers, purchasers, providers, policy makers) can understand the results of the measure and are likely to find them useful for decision making. (composite measure evaluation criteria)

Eval

3a. Meaningful, Understandable, and Useful Information

Current Use: ☒ In use ☐ Not in use, but testing completed ☐ Testing not yet completed

If used in a public reporting initiative, Name of initiative(s), locations, Web page URL(s): N/A

If used in other programs/initiatives (e.g., quality improvement), Name of initiative(s), locations, web page URL(s): N/A

Testing of Interpretability (*Testing that demonstrates the results are understood by the potential users for public reporting and quality improvement*)

►**Data/sample** *(description of data/sample and size)*: No formal testing has been performed. However, provider feedback has been uniformly positive, and multiple payers have been given the opportunity to comment on the

3a
H☒ M☐ L☐ N☐

usability of the measure. Their comments have been universally positive, and no substantive recommendations for change were made.

►**Methods** *(methods, e.g., focus group, survey, QI project)*:

►**Results** *(qualitative and/or quantitative results and conclusions)*:

3b/3c. Relation to other NQF-endorsed measures
Identify similar or related NQF-endorsed measures (available at *www.qualityforum.org under Core Documents)*

☐ Other measures for same target population ☒ Other measures on same topic ☐ No similar measures

NQF # and Title of similar or related measures:
NQF # 0127 - Pre-Operative Beta Blockade
NQF # 0117 - Beta Blockade at Discharge
NQF # 0116 - Anti-Platelet Medication at Discharge
NQF # 0118 - Anti-Lipid Treatment Discharge
NQF # 0134 - Coronary artery bypass graft (CABG) using internal mammary artery (IMA)
NQF # 0119 - Risk-Adjusted Operative Mortality for CABG©
NQF # 0131 - Stroke/Cerebrovascular Accident
NQF # 0115 - Surgical Re-exploration
NQF # 0130 - Deep Sternal Wound Infection Rate
NQF # 0114 - Post-operative Renal Failure
NQF # 0129 - Prolonged Intubation (ventilation)

Describe the distinctive or additive value this measure provides to existing NQF-endorsed measures:
Combining these 11 individual measures into a composite provides a much more comprehensive and multidimensional assessment of CABG quality while presenting it as one easily understood score. Furthermore, by combining multiple individual endpoints, our ability to discriminate levels of performance among providers is enhanced compared to using just one endpoint such as mortality.

3b. Harmonization

►**Are the component measure specifications harmonized, or if not, why?** Yes.

3b
H☒
M☐
L☐
N☐
NA☐

3c. Distinctive or Additive Value

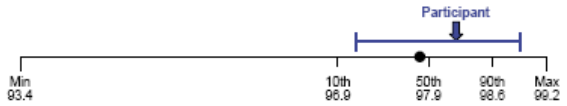
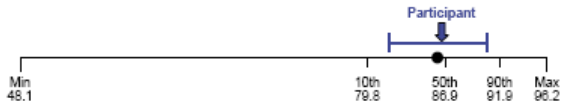
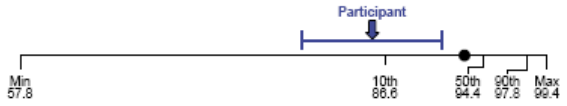
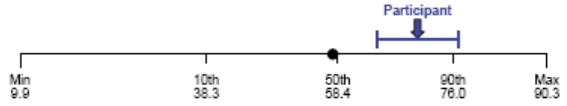
►**Describe the distinctive, improved, or additive value this measure provides to existing NQF-endorsed measures:**
Combining these 11 individual measures into a composite provides a much more comprehensive and multidimensional assessment of CABG quality while presenting it as one easily understood score. Furthermore, by combining multiple individual endpoints, our ability to discriminate levels of performance among providers is enhanced compared to using just one endpoint such as mortality.

3c
H☒
M☐
L☐
N☐
NA☐

3d. Decomposition of Composite
►**Describe the information from decomposing the composite into its components that is available:**
Since the inception of the STS CABG composite, we have considered it essential that the overall score be decomposable into its component domains and measures. In each of the five semi-annual reports we have provided thus far, STS participants are given their percentile score (with confidence intervals), the STS mean and median, and their "star rating" (one-star = 99% Bayesian probability that the provider's performance is below that of the STS average; two-star = statistically indistinguishable from the STS average; three-star = 99% Bayesian probability that the provider's performance exceeds that of the STS average). The same scoring is provided for each individual domain. This is illustrated by the following report example:

Page 1: Scores and Star Ratings Sample Results

3d
H☒
M☐
L☐
N☐

Quality Domain	Participant Score (98% CI)	STS Mean Participant Score	Participant Rating[1]	Distribution of Participant Scores • = STS Mean
2006 Overall	95.3% (94.1 , 96.3)	94.5%	★★	
2006 Avoidance of Mortality	98.2% (97.1 , 98.9)	97.8%	★★	
2006 Avoidance of Morbidity[2]	86.6% (81.8 , 90.7)	86.2%	★★	
2006 Use of IMA[3]	85.6% (80.0 , 91.1)	92.9%	★	
2006 Medications[4]	70.6% (64.3 , 76.7)	57.6%	★★★	

Finally, for the two any-or-none or all-or-none domains, we provide a listing of how often the domain was a "failure" because of each individual component measure. This enables providers to focus their performance improvement activities. An example of this listing is as follows:

Page 2:Quality Domain Details Sample Results

Quality Domain	Eligible Procedures	Detail	Count	Percent of Morbidity/Failure[1]
2006 Avoidance of Mortality	553	Mortality	6	
2006 Avoidance of Morbidity[2]	553	Any Morbidity	73	
		Reoperation only[3]	14	19.2%
		Renal Failure only[4]	8	11.0%
		Deep Sternal Wound Infection only	0	0.0 %
		Prolonged Ventilation only	33	45.2%
		Cerebrovascular Accident only[5]	1	1.4 %
		Multiple Morbidities	17	23.3%
2006 Use of IMA[6]	522	IMA Failures	75	
2006 Medications[7]	553	Failed to Prescribe all eligible NQF Medications	158	
		Only failed to prescribe Preoperative Beta Blockade	70	44.3%
		Only failed to prescribe Discharge Beta Blockade[8]	17	10.8%
		Only failed to prescribe Discharge Anti-Lipids[8]	31	19.6%
		Only failed to prescribe Discharge Anti-Platelets[9]	16	10.1%
		Failed to prescribe multiple medications	24	15.2%

Page 2: Quality Domain Details - Sample Results Interpretation

Avoidance of Mortality: Out of the 553 procedures eligible for this measure, 6 died corresponding to a 1.1% observed mortality rate. Note that while this number represents the observed mortality, the participant score for this measure (from page 1) is risk-adjusted and came from multivariable hierarchical analysis. For this reason a direct correlation between these sets of numbers should not be expected.

Avoidance of Morbidity (All-or-None measure): Of the 553 procedures eligible for this measures, 73 (13.2%) had at least one of the morbidities for this all-or-none measure. Of the 73 that had at least one of the morbidities, 14 (19.2%) had reoperation only, 8 (11.0%) had renal failure only, 0 had prolonged ventilation only, 33 (45.2%) had cerebrovascular accident only and 17 (23.3%) actually had more than one of the morbidities. Of the cases that had a single morbidity, prolonged ventilation appears to account for the highest proportion. This detailed information is designed to allow database participants to understand what specific quality measures are driving their quality scores for the four domains.

Use of IMA: Of the 522 procedures eligible for this measure, 75 (14.4%) did not receive an IMA. Alternatively, 85.6% did receive an IMA. Note that for these sample results this number directly correlates with the 85.6% participant score for this measure on page 1 of this sample. However, these numbers may not always correlate because the quality scores are estimates of true performance based upon hierarchical analysis. See section IV of the General Report Overview for more information about why performance estimation results from hierarchical analysis do not always match observed results.

Medications (All-or-None Measure): Of the 553 procedures eligible for these measures, 158 (28.6%) did not receive all the eligible NQF medications. Of these 158 medication failures, preoperative beta blockade therapy was the only medication failure in 70 (44.3%) of them. Discharge beta blockade therapy was the only medication failure in 17 (10.8%), discharge anti-lipid medication was the only medication failure in 31 (19.6%), and discharge anti-platelet medication was the only medication failure in 16 (10.1%). Multiple medication failures accounted for 24 (15.2%) of the 158 cases.

The largest proportion of cases with medication failures had them for preoperative beta blockade therapy alone. In those cases where the largest proportion of cases with failures is for multiple medications, the participant will need to explore in more detail to determine which specific combinations of medications were not prescribed. Note that a direct comparison between these domain detail results and those in the NQF section of the report is not possible because a) the denominator is cases with any medication failure (158) and b) the results for an individual medication (i.e. preoperative beta blockade therapy) show only those cases that missed ONLY this medication. Other cases that missed additional medications in addition to preoperative beta blockade therapy would be counted under the multiple medication failure item.

3e. Achieved stated purpose
Describe how the results reported above demonstrate that the composite achieves the stated purpose: Based on the measures and domains included in this composite, its usability, and the results of pilot testing, we believe the STS CABG composite achieves our stated goal: to develop a comprehensive, multidimensional, composite score for CABG performance.

3e
H☒
M☐
L☐
N☐

Staff Notes to Reviewers (including additions/changes to related or similar measures): NQF #0076 CAD: optimally managed modifiable risk

Steering Committee/TAP: Overall, to what extent was the criterion, *Usability*, met?
Rationale: in use; uses clinical data; data provided to providers; may need public education to understand composites in general;

3
H☒
M☐
L☐
N☐

4. FEASIBILITY

Extent to which the required data are readily available, retrievable without undue burden, and can be implemented for performance measurement. (composite measure evaluation criteria)

Eval

4a. Data Generated as a Byproduct of Care Processes

How are all the data elements that are needed to compute measure scores generated? *Check all that apply*
☒ Data are generated as a byproduct of care processes during care delivery (*Data are generated and used by*

4a
H☒
M☐
L☐

healthcare personnel during the provision of care, e.g., blood pressure, lab value, medical condition)
☒ Coding/abstraction performed by someone other than person obtaining original information *(e.g., DRG, ICD-9 codes on claims; chart abstraction for quality measure, registry)*
☒ Other *(e.g., patient experience of care surveys, provider surveys, observation)*, *Please describe*:

N☐
NA☐

4b. Electronic Sources

►**Are all the data elements available electronically?** (*elements that are needed to compute measure scores are in defined, computer-readable fields, e.g., electronic health record, electronic claims*)
☒ Yes ☐ No
►**If no, specify the near-term path to achieve electronic capture by most providers.**

Note: Measure stewards will be asked to specify the data elements for electronic health records at a later date

4b
H☒
M☐
L☐
N☐

4d. Susceptibility to Inaccuracies, Errors, or Unintended Consequences

►**Identify susceptibility to inaccuracies, errors, or unintended consequences of the measure and describe how these potential problems could be audited. If audited, provide results.**
All public reporting initiatives have the potential for unintended consequences, including gaming and risk aversion. We attempt to control the former through a careful audit process, and the latter by having a robust methodology that appropriately adjusts the expected risk for providers who care for sicker patients.

4d
H☒
M☐
L☐
N☐

4e. Data Collection Strategy/Implementation

►**Describe what you have learned/modified as a result of testing and/or operational use of the composite/component measures regarding data collection, availability of data/missing data, timing/frequency of data collection, patient confidentiality, time/cost of data collection, other feasibility/ implementation issues:**
No significant issues have been discovered during 2 ½ years of real-world implementation, including about 90% of cardiac surgery providers in the US. Timely ascertainment of thirty-day mortality remains a challenge for some programs (not a problem unique to our registry), and we have addressed this by a number of approaches: (1) enhanced audit, with special emphasis on 30-day status; (2) verification and supplementation of provider-submitted mortality data with external sources including the Social Security Death Master File; (3) the ability to provide two very closely related CABG composite scores, one using operative-mortality (30-day regardless of venue, in-hospital regardless of timing) and one using in-hospital mortality (which we capture with nearly 100% accuracy based on validation with CMS data). Initial studies suggest that the results of these two approaches are quite similar

►**Costs to implement the measure** (*costs of data collection, fees associated with proprietary measures*)**:**
Data Collection:
The composite measure is a combination of 11 measurers that are routinely collated with surgical procedures in the Adult Cardiac Surgery Database. Consequently there are no direct costs to collect the data for the composite quality measure.

Costs to develop the measure included volunteer cardiothoracic time, STS staff time, and DCRI statistician and project management time.

Fees Associated with Proprietary Measures:
STS Adult Cardiac Surgery Database participants (single cardiothoracic surgeons or a group of surgeons pay annual participant fees of $2,750 or $3,450 depending on whether participants in which the majority of surgeons are STS members pay the lower fees as a benefit of membership. The composite measure is one of many measures that are reported to participants. Therefore, the costs of these measures cannot be disaggregated from overall report costs.

Third parties that request composite or other measures pay fees that are dependent on the data requested. STS does not provide any data with out explicit written consent
►**Evidence for costs:**
►**Business case documentation:**

4e
H☒
M☐
L☐
N☐

Staff Notes to Reviewers:

Reviewers: Overall, to what extent was the criterion, Feasibility, met? **Rationale:** uses clinical database; audited	**4** **H**☒ **M**☐ **L**☐ **N**☐
Reviewers: Overall, to what extent were all the criteria met? **Rationale:**	**H**☐ **M**☐ **L**☐
Steering Committee only **Recommendation:** ☐ **Endorsement** ☐ **Time-limited endorsement** ☐ **Do not recommend** **Conditions:** ☐ **No** ☐ **Yes, Specify:**	

CONTACT INFORMATION

Measure Steward (Intellectual Property Owner)
Organization: The Society of Thoracic Surgeons
Street Address: 633 N. Saint Clair St., Suite 2320 **City:** Chicago **State:** IL **ZIP:** 60611

Point of Contact: First Name: Jane **MI:** M **Last Name:** Han **Credentials (MD, MPH, etc.):** MSW
Email: jhan@sts.org **Telephone:** 312-202-5856 ext:

Measure Developer If different from Measure Steward
Organization:
Street Address: **City:** **State:** **ZIP:**

Point of Contact: First Name: **MI:** **Last Name:** **Credentials (MD, MPH, etc.):**
Email: **Telephone:** ext:

Submitter If different from Measure Steward Point of Contact
First Name: **MI:** **Last Name:** **Credentials (MD, MPH, etc.):**
Email: **Telephone:** ext:
Organization: ☐ Measure Steward ☐ Measure Developer

Additional Measure Developer Organizations: 0

ADDITIONAL INFORMATION

Workgroup/Expert Panel involved in measure development
►Provide a list of sponsoring organizations and workgroup/panel members' names and organizations.

STS TASK FORCE ON QUALITY MEASUREMENT (2007)

Member Name	Institution	Location
David M. Shahian, MD - Chair	Massachusetts General Hospital	Boston, MA
Richard P. Anderson, MD*	---	Seattle, WA
Elizabeth DeLong, PhD	Duke Clinical Research Institute	Durham, NC
Rachel S. Dokholyan, MPH	Duke Clinical Research Institute	Durham, NC
Fred H. Edwards, MD	University of Florida, Shands Jacksonville	Jacksonville, FL
Victor A. Ferraris, MD, PhD	University of Kentucky College of Medicine	Lexington, KY
Constance K. Haan, MD, MS	University of Florida College of Medicine- Jacksonville	Jacksonville, FL
Sharon-Lise T. Normand, PhD	Harvard Medical School Department of Health Care Policy	Boston, MA
Sean M. O'Brien, PhD	Duke Clinical Research Institute	Durham, NC
Eric D. Peterson, MD, MPH	Duke Clinical Research Institute	Durham, NC

Jeffrey B. Rich, MD	Sentara Heart Hospital	Norfolk, VA
Cynthia M. Shewan, PhD	The Society of Thoracic Surgeons	Chicago, IL

* This member is deceased

►Describe the members' role in measure development.
The 2007 STS Task Force on Quality Measurement members collectively formulated the composite measure goals and principles, participated in the methodological development and testing, and continue to be involved in its implementation.

►**If adapted, provide name of original measure:**
►**If adapted, provide original specifications ☐ attachment or web page URL:**

Measure Developer/Steward Updates and Ongoing Maintenance
►Year the measure was first released: 2007
►Month and Year of most recent revision: April 2007
►What is the frequency for review/update of this measure? Every three years
►When is the next scheduled review/update for this measure? 2010

Copyright statement/disclaimers:

Additional Information web page URL:
Attachment 1- http://ats.ctsnetjournals.org/cgi/content/abstract/88/1_Supplement/S2
Attachment 2- http://ats.ctsnetjournals.org/cgi/content/full/83/4_Supplement/S3
Attachment 3- http://ats.ctsnetjournals.org/cgi/content/full/83/4_Supplement/S13

I have checked that the submission is complete and all the information needed to evaluate the measure is provided in the form; any blank fields indicate that no information is provided.☒

Date of Submission (*MM/DD/YY*)**:** 9/18/09

THE SOCIETY OF THORACIC SURGEONS

633 N. SAINT CLAIR STREET, SUITE 2320
CHICAGO, ILLINOIS 60611-3658
Phone: 312/202-5800
Fax: 312/202-5801
E-mail: sts@sts.org
Web: http://www.sts.org

April 26, 2010

The National Quality Forum
601 Thirteenth Street NW
Suite 500 North
Washington, DC 20005

Re: NQF Patient Outcomes Steering Committee Discussion on April 20, 2010 – STS Response

Dear Steering Committee Members:

During your meeting on April 20, 2010, a number of concerns were expressed regarding The Society of Thoracic Surgeons (STS) CABG Composite Score (OT1-013-09). Unfortunately, time constraints made it impossible for STS to respond to each concern in detail. Thus, on behalf of STS, we would like to take this opportunity to provide responses to the main questions raised regarding the STS composite measure. Please see below:

1. Concern that exclusion and inclusion criteria for NQF-endorsed component measures of the STS composite were not provided

   STS Response:
   The exclusion and inclusion criteria are included in each composite report to STS Adult Cardiac Surgery Database participants. The table describing these criteria, taken directly from our standard report, is provided below in Appendix A.

2. Dissatisfaction with the 99% Bayesian certainty criterion for star rating was expressed

   STS Response:
   Numerical performance scores, including point estimates and confidence intervals for the participant as well as overall STS scores and percentiles, are routinely calculated and provided in STS' standard reports. Unfortunately, the vast majority of consumers would not understand how to correctly interpret these data. It was for this reason that we designed the one, two and three star rating system. We tested various Bayesian probabilities to determine one and three star rating categories, ultimately deciding on 99%. This strict criterion assures that programs designated as one- or three-star have unequivocally different performance from the STS average. On the other hand, it has consistently produced about 10-15% one-star and 10-15% three-star programs each harvest period. This is an order of magnitude higher number of low and high outlier programs than could be identified using risk-adjusted mortality alone, and it is also a far greater percentage of outliers than identified in any credible CABG public reporting system of which we are aware. Finally, as shown in Appendix B, the star ratings (which

correspond to low, mid, and high in the Figure) correlate well with actual clinical performance in each of the component domains of the composite.

We believe the currently constructed star system is clinically meaningful; it protects providers from spurious identification as low outliers; it prevents programs from being designated as three-star unless they are truly superior; it correctly identifies more programs as low or high outliers than any other system of which we are aware; and it provides a readily comprehensible single rating for consumers.

We believe that it would be unwise to endorse our overall methodology and numerical score but not our star rating, which is an integral part of our methodology. NQF may not have endorsed such a comprehensive approach to both measurement and performance rating before, but we believe this is a logical evolutionary step for NQF and STS, particularly given our long history of data collection and performance measurement. Finally, we are concerned that publishing only numerical scores could have unintended negative consequences. Without our consent or input, external entities could construct their own performance rating systems that would not be justified by the data underlying our scores.

3. Concern regarding temporal shifts in star ratings

STS Response:
One member of the Committee was concerned that about 50% of one- and three-star programs changed their star ratings over the course of a year. The information on which that concern is based is found in Appendix C.

First, our 99% Bayesian probability criterion is strict, as noted above. One must truly be statistically quite different from the average STS performer to receive one or three stars. The fact that many programs move from the one- and three-star categories to the average two-star category is not unexpected—small differences in performance over the course of a year may account for this, especially for providers that are right on the borderline between two-star and either one- or three-star performance. A much smaller percentage of programs move from being average to being one or three stars, at less than 10% each. Notably, virtually no providers move from one to three or three to one stars. Such dramatic short-term changes in performance rating would be of concern, suggesting instability of our methodology.

4. Concern that consumers would misinterpret one, two, and three stars as being *Good, Better*, and *Best*, respectively

STS Response:
As exemplified in Appendix D, which was also included in our submission, we have always indicated that one-star programs are performing below the STS average. The key to avoiding consumer misinterpretation of this or any other performance rating system is to provide clear explanations. In our discussions with Consumers Union regarding our collaborative public reporting initiative, both parties have agreed completely on the importance of extensive educational content to clarify the correct interpretation of our rating system.

Please do not hesitate to contact Jane Han, STS Manager of Quality Initiatives, at jhan@sts.org or (312) 202-5856, with any questions you may have. Thank you for your time and consideration.

Sincerely,

David M. Shahian, MD
Chair, STS Workforce on National Databases
Chair, STS Quality Measurement Task Force

Fred H. Edwards, MD
Chair, STS Council on Quality, Research, and Patient Safety Operating Board

Frederick L. Grover, MD
Immediate Past Chair, STS Council on Quality, Research, and Patient Safety Operating Board

cc: Reva Winkler, MD, MPH, Program Consultant
Heidi Bossley, MSN, MBA, Senior Director, Performance Measures
Hawa Camara, Research Analyst
Sarah Fanta, Research Analyst, Performance Measures

Appendix A

Report Overview
STS Composite Quality Rating and NQF Measures
STS Report – Period Ending 06/30/2009

Table 2. STS Implementation of the NQF Measures – Updated 4/3/2009

MEASURE	STS implementation	Comments
1. Participation in a Systematic Database for Cardiac Surgery *"Does the facility participate in a multicenter data collection and feedback program that provides benchmarking relative to peers and uses process and outcome measures?"*	Not reported.	NOTE: All report recipients participate in a systematic database for cardiac surgery (STS).
2. Surgical Volume: a. Isolated CABG b. Valve Surgery c. Valve+CABG Surgery *"Annual procedural volume of three surgeries: isolated CABG surgery, valve surgery, and valve+CABG surgery"*	a. Isolated CABG (Same population definition as in the STS harvest report – see Table 9 of the Report Overview). Variables used: CABG (OpCab)[1] b. Valve Surgery – Any mitral, aortic, tricuspid, or pulmonary valve surgery without a CABG. Variables used: Mitral valve surgery (OpMitral), Aortic valve surgery (OpAortic), Triscupid valve surgery (OpTricus), Pulmonary valve surgery (OpPulm) c. CABG + Valve Surgery – Any mitral, aortic, tricuspid, or pulmonary valve surgery with a CABG Variables used: Mitral valve surgery (OpMitral), Aortic valve surgery (OpAortic), Triscupid valve surgery (OpTricus), Pulmonary valve surgery (OpPulm), and CABG (OpCab)	NOTE: NQF procedure groups determined by ICD-9 code. STS does not collect ICD-9 codes.
3. Timing of Antibiotic Administration for Cardiac Surgery Patients *"Percent of patients undergoing cardiac surgery who received prophylactic antibiotics within one hour of surgical incision (two hours if receiving vancomycin)."*	Not reported	NOTE: STS began collecting information on antibiotic administration with data version 2.61 but will not report on this measure until at least 2009.

Report Overview
STS Composite Quality Rating and NQF Measures
STS Report – Period Ending 06/30/2009

4. Selection of Antibiotic Administration for Cardiac Surgery Patients *"Percent of patients undergoing cardiac surgery who received prophylactic antibiotics recommended for the operation."*	Not reported	NOTE: STS began collecting information on antibiotic administration with data version 2.61 but will not report on this measure until at least 2009.
5. Pre-operative Beta Blockade *"Percent of patients undergoing isolated CABG who received beta blockers within 24 hours preceding surgery."*	• Procedure: Isolated CABG • Variable used: Meds-Beta Blockers (MedBeta) • Numerator: Number of Isolated CABG procedures in which (MedBeta) is marked as 'Yes' • Denominator: Total number of Isolated CABG procedures excluding those in which (MedBeta) is marked as 'Contraindicated/Not indicated'	NOTE: STS began collecting information on whether medications were contraindicated/not indicated with data version 2.61. Beginning with 2008 harvest 3 these cases are removed from the denominator.
6. Use of Internal Mammary Artery (IMA). *"Percent of patients undergoing isolated CABG who received an IMA graft."*	• Procedure: Isolated CABG • Variable: IMA Artery Used (IMAArtUs) • Numerator: Number of Isolated CABG procedures in which (IMAArtUs) is marked as 'Left IMA', 'Right IMA', or 'Both IMAs' • Denominator: Number Isolated CABG excluding repeat CABG (PrCAB)	NOTE: NQF population definition and exclusions are based on ICD-9 codes. STS does not currently collect ICD-9 codes. NOTE: The NQF exclusion of other heart procedures is obtained during STS implementation by definition of the isolated CABG group (See Table 9 of the Report Overview).
7. Duration of Prophylaxis for Cardiac Surgery Patients *"Percent of patients undergoing cardiac surgery who prophylactic antibiotics were discontinued within 24 hours after surgery end time."*	Not reported	NOTE: STS does not currently collect information on antibiotic administration.

Appendix A

Report Overview
STS Composite Quality Rating and NQF Measures
STS Report – Period Ending 06/30/2009

8. Prolonged Intubation (ventilation). *"Percent of patients undergoing isolated CABG (without pre-existing intubation/tracheostomy) who require intubation for more than 24 hours."*	• Procedure: Isolated CABG • Variable: Complications-Pulmonary_Vent Prolonged (CPVntLng) • Numerator: Number of Isolated CABG procedures in which (CPVntLng) is marked as 'Yes' • Denominator: Total number of Isolated CABG procedures • Risk adjustment: Yes[2]	NOTE: STS does not collect data on the NQF exclusion of intubation/tracheostomy prior to isolated CABG.
9. Deep Sternal Wound Infection Rate. *"Percent of patients undergoing isolated CABG who developed deep sternal wound infection within 30 days post-operatively."*	• Procedure: Isolated CABG • Variable: Complications–Infection-Sternum-Deep (CIStDeep) • Numerator: Number of Isolated CABG procedures in which (CIStDeep) is marked as 'Yes' • Denominator: Total number of Isolated CABG procedure Risk adjustment: Yes	NOTE: Through data version 2.52.1 Deep Sternal Wound Infection Rate was only tracked up to discharge. Beginning with data version 2.61 this rate is being tracked for 30 days postoperatively. NOTE: STS does not currently collect information on pre-operative wound site infections and cannot apply the NQF exclusion.
10. Stroke/Cerebrovascular Accident. *"Percent of patients undergoing isolated CABG (without pre-existing neurologic deficit) who develop a post-operative neurologic deficit persisting greater than 72 hours."*	• Procedure: Isolated CABG • Variable: Complications – Neurologic-Stroke-Permanent (CNStrokP) • Numerator: Number of Isolated CABG procedures in which (CNStrokP) is marked as 'Yes' • Denominator: Number of Isolated CABG procedures excluding those with a prior CVA (CVA) • Risk adjustment: Yes	NOTE: STS implementation excludes patients with prior CVA. NQF has an exclusion for "neurologic deficits" that is not explicitly defined. NOTE: Beginning with data version 2.61, the STS definition includes deficits persisting greater than 24 hours.
11. Post-operative Renal Insufficiency. *"Percent of patients undergoing isolated CABG (without pre-existing renal failure) who develop post-operative renal failure or require dialysis."*	• Procedure: Isolated CABG • Variable: Complications-Renal_Renal Failure (CRenFail) • Numerator: Number of Isolated CABG procedures in which (CRenFail) is marked as 'Yes'. • Denominator: Number of Isolated CABG procedures excluding those with either or both of the following risk factors: Renal Failure (RenFail) for data version 2.52.1 or earlier, Dialysis (Dialysis) for data version 2.61 or later, Last Creatinine Level (CreatLst) > 2 • Risk adjustment: Yes	NOTE: Although both NQF and STS refer to the same underlying clinical definition, the NQF uses the label "renal insufficiency" and the STS uses the label "renal failure". For the purposes of this report the STS has labeled this measure "Post-operative renal insufficiency (failure)"
12. Surgical Re-exploration. *"Percent of patients undergoing isolated CABG who require a return to the operating room for bleeding/tamponade, graft occlusion, or other cardiac reason."*	• Procedure: Isolated CABG • Variables: Complications-Operative-ReOperation for Bleeding/Tamponade (COpReBld), Complications-Operative-ReOperation for Graft Occlusion (COpReGft), Complications-Operative-ReOperation for Other Cardiac Reasons (COpReOth), Complications-Operative-ReOperation for Valve Dysfunction (COpReVlv) • Numerator: Number of Isolated CABG procedures in which any of the variables above are marked 'Yes' • Denominator: Total number of Isolated CABG procedures • Risk adjustment: Yes	
13. Anti-platelet Medications at Discharge. *"Percent of patients undergoing isolated CABG who were discharged on aspirin/safety-coated aspirin or clopidogrel."*	• Procedure: Isolated CABG • Variables: For data version 2.41: Discharge Medications-Aspirin (DCASA) or Discharge Medications-Other Anti-platelets (DCAntplt) For data versions 2.52.1 and 2.61: Discharge Medications-Aspirin (DCASA) or Discharge Medications-ADP Inhibitors (DCADP) • Numerator: Number of Isolated CABG procedures in which any of the above variables are marked 'Yes' • Denominator: Number of Isolated CABG procedures excluding those that were submitted under STS data version 2.35, those that resulted in in-hospital mortalities based on the variables Mortality Discharge Status (MtDCStat), Mortality Date (MtDate), and Discharge Date (DischDt), and those submitted under STS data version 2.61 in which (DCASA or DCADP) is marked as 'Contraindicated/Not indicated'	NOTE: Although the NQF measure does not exclude patients who died in the hospital, the STS implementation does exclude in-hospital mortalities. NOTE: STS implementation excludes records collected under v2.35 when information on discharge anti-platelet medications was not collected. NOTE: STS began collecting information on whether medications were contraindicated/not indicated with data version 2.61. Beginning with 2008 harvest 3 these cases are removed from

Report Overview
STS Composite Quality Rating and NQF Measures
STS Report – Period Ending 06/30/2009

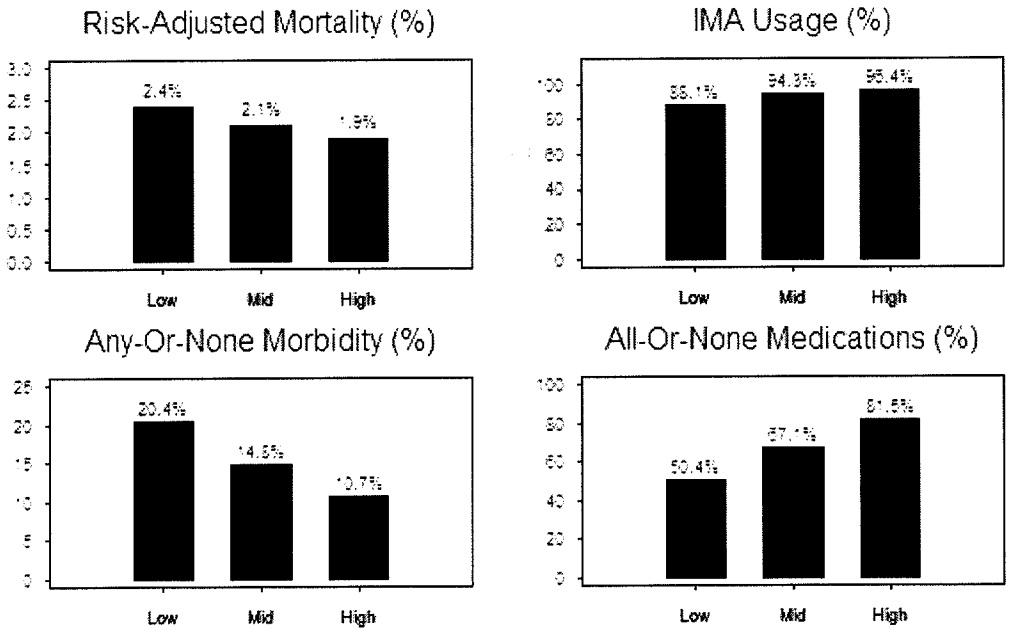
Report Overview
STS Composite Quality Rating and NQF Measures
STS Report – Period Ending 06/30/2009

		the denominator.
14. Beta Blockade at Discharge *"Percent of patients undergoing isolated CABG who were discharged on beta blockers."*	• Procedure: Isolated CABG • Variable: Discharge Medications-Beta Blockers (DCBeta) • Numerator: Number of Isolated CABG procedures in which (DCBeta) is marked 'Yes' • Denominator: Number of Isolated CABG procedures excluding those that were submitted under STS data version 2.35, those that resulted in in-hospital mortalities based on Mortality Discharge Status • (MtDCStat) and Mortality Date (MtDate), and those submitted under STS data version 2.61 in which (DCBeta) is marked as 'Contraindicated/Not indicated'	NOTE: Although the NQF measure does not exclude patients who died in the hospital, the STS implementation does exclude in-hospital mortalities. NOTE: STS implementation excludes records collected under v2.35 when information on discharge beta blockade was not collected. NOTE: STS began collecting information on whether medications were contraindicated/not indicated with data version 2.61. Beginning with 2008 harvest 3 these cases are removed from the denominator.
15. Anti-lipid Treatment at Discharge *"Percent of patients undergoing isolated CABG who were discharged on a statin or other pharmacologic lipid-lowering regimen."*	• Procedure: Isolated CABG • Variables: Discharge Medications-Lipid Lowering (DCLipid) • Numerator: Number of Isolated CABG procedures in which (DCLipid) is marked as 'Yes' • Denominator: Number of Isolated CABG procedures excluding those that were submitted under STS data version 2.35, those that resulted in in-hospital mortalities based on Mortality Discharge Status (MtDCStat), Mortality Date (MtDate), and Discharge Date (DischDt), and those submitted under STS data version 2.61 in which (DCLipid) is marked as 'Contraindicated/Not indicated'	NOTE: Although the NQF measure does not exclude patients who died in the hospital, the STS implementation does exclude in-hospital mortalities. NOTE: STS implementation excludes records collected under v2.35 when information on discharge anti-lipid medication was not collected. NOTE: STS began collecting information on whether medications were contraindicated/not indicated with data version 2.61. Beginning with 2008 harvest 3 these cases are removed from the denominator.
16. Risk-Adjusted Inpatient Operative Mortality for CABG.	• Population: Isolated CABG • Variable: In-hospital mortalities based on Mortality Discharge	NOTE: NQF population currently defined by CCMRP; STS population defined by

Report Overview
STS Composite Quality Rating and NQF Measures
STS Report – Period Ending 06/30/2009

"Percent of patients who die in hospital after CABG surgery."	Status (MtDCStat) and Mortality Date (MtDate) • Risk adjustment: Yes • Numerator: Number of Isolated CABG procedures with an in-hospital mortality • Denominator: Total number of Isolated CABG procedures	STS procedure groups (isolated CABG) and may not match CCMRP exclusions exactly.
17. Risk-Adjusted Operative Mortality for CABG *"Percent of patients undergoing isolated CABG who die, including both 1) all deaths occurring during the hospitalization in which the CABG was performed, even if after 30 days, and 2) those deaths occurring after discharge from the hospital, but within 30 days of the procedure."*	• Procedure: Isolated CABG • Variables: Operative mortality based on Mortality Operative Death (MtOpD), Mortality Status at 30 days (Mt30Stat), Mortality Date (MtDate), and Mortality Discharge Status (MtDCStat) • Numerator: Number of Isolated CABG procedures with an operative mortality • Denominator: Total number of Isolated CABG procedures • Risk adjustment: Yes	
18. Risk-Adjusted Operative Mortality for Aortic Valve Replacement (AVR) *"Percent of patients undergoing AVR who die, including both 1) all deaths occurring during the hospitalization in which the [procedure] was performed, even if after 30 days, and 2) those deaths occurring after discharge from the hospital, but within 30 days of the procedure."*	• Procedure: Isolated AV Replacement (Same population definition as in the STS harvest report – see Table 9 of the Report Overview) • Variables: Operative mortality based on Mortality Operative Death (MtOpD), Mortality Status at 30 days (Mt30Stat), Mortality Date (MtDate), and Mortality Discharge Status (MtDCStat) • Numerator: Number of Isolated AV Replacement procedures with an operative mortality • Denominator: Total number of Isolated AV Replacement procedures • Risk adjustment: Yes	
19. Risk-Adjusted Operative Mortality for Mitral Valve Replacement/Repair (MVR) *"Percent of patients undergoing MVR who die, including both 1) all deaths occurring during the hospitalization in which the [procedures] was performed, even if after 30 days, and 2) those deaths occurring after discharge from the hospital,*	• Procedure: Isolated MV Replacement (Same population definition as in the STS harvest report – see Table 9 of the Report Overview) • Variables: Operative mortality based on Mortality Operative Death (MtOpD), Mortality Status at 30 days (Mt30Stat), Mortality Date (MtDate), and Mortality Discharge Status (MtDCStat) • Numerator: Number of Isolated MV Replacement procedures	NOTE: Although the NQF lists the STS as the source for this measure, their population definition does not match current STS population definitions. STS implementation excludes MV repair patients because STS risk adjustment only exists

Appendix B


Risk-Adjusted Mortality (%)
2.4%
2.1%
1.9%
Low
Mid
High
IMA Usage (%)
88.1%
94.3%
95.4%
Low
Mid
High
Any-Or-None Morbidity (%)
20.4%
14.5%
10.7%
Low
Mid
High
All-Or-None Medications (%)
50.4%
67.1%
81.5%
Low
Mid
High

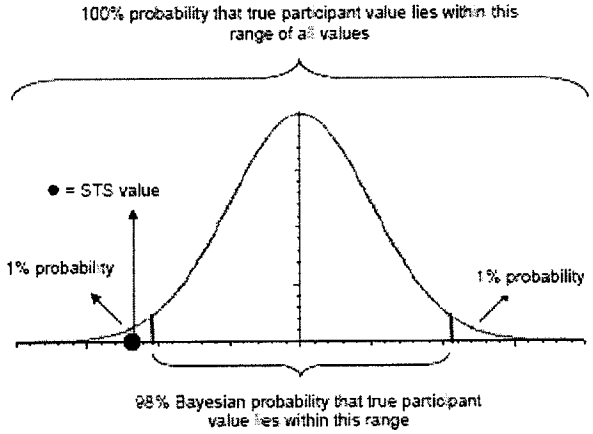
Appendix C

Table 3. Change in star ratings between 1st and 4th harvests periods

Spring 2007	Fall 2008		
	1	2	3
	N	N	N
1	44	44	1
2	31	434	46
3	1	51	54

Appendix D

Report Overview
STS Composite Quality Rating and NQF Measures
STS Report – Period Ending 06/30/2008



Column 3. STS Mean Participant Score. The STS mean participant score is the average of all scores across all of the participants in the analysis. This score serves as a useful benchmark for assessing a participant's performance relative to the overall STS performance.

Column 4. Participant Rating. The participant rating system assigns participants to rating categories designated by one, two, or three stars. The rating categories are defined as follows:

★★★ → Participant performance is significantly higher than STS mean.

★★ → Participant performance is not statistically different from STS mean.

★ → Participant performance is significantly lower than STS mean.