NATIONAL QUALITY FORUM

MEASURE WORKSHEET

This document summarizes the evaluation of the measure as it progresses through NQF's Consensus Development Process (CDP). The information submitted by measure developers/stewards is included after the Brief Measure Information, Preliminary Analysis, and Pre-meeting Public and Member Comments sections.

To navigate the links in the worksheet: Ctrl + click link to go to the link; ALT + LEFT ARROW to return

Purple text represents the responses from measure developers.

Red text denotes developer information that has changed since the last measure evaluation review.

Brief Measure Information

NQF #: 0117

Corresponding Measures:

De.2. Measure Title: Beta Blockade at Discharge

Co.1.1. Measure Steward: The Society of Thoracic Surgeons

De.3. Brief Description of Measure: Percent of patients aged 18 years and older undergoing isolated CABG who were discharged on beta blockers

1b.1. Developer Rationale: The use of postoperative b-blockers is now known to protect patients both at one year and long term (greater than 5 years) from death following cardiac surgery. This effect is associated with a 46 % risk reduction in death at one –year and 35% risk reduction in mortality during long-term follow-up (see Chan below). The summary of peer reviewed literature cited below supports that the utilization of beta-blocker at discharge as conferring a strong risk reduction in mortality.

- Crystal E, Connolly SJ, Sleik K, et al. Interventions on prevention of postoperative atrial fibrillation in patients undergoing heart surgery: a meta-analysis. Circulation. 2002;106(1):75-80.
- 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.
- Maisel WH, Rawn JD, Stevenson WG. Atrial fibrillation after cardiac surgery. Ann Intern Med. 2001;135(12):1061-1073.
- 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.
- Welke KF, Ferguson TB, Coombs LP, et al. Validity of the Society of Thoracic Surgeons National Adult Cardiac Surgery Database. Ann Thorac Surg. 2004; 77:1137-1139.
- Charlson ME, Isom OW. Care after coronary-artery bypass surgery. N Engl J Med. 2003; 348:1456-63.

- Chen J, Radford MJ, Wang Y, Marciniak TA, Krumholz HM. Are beta-blockers effective in elderly patients who undergo coronary revascularization after acute myocardial infarction? Arch Intern Med. 2000; 160:947-52.
- Chan AYM, McAlister FA, Norris, CM, et al. Effect of B-Blocker use on outcomes after discharge in patients who underwent cardiac surgery. J Thorac Cardiovasc Surg. 2010; 140:182-7.
- Zhang H, Yuan X, Zhang H, et al. Efficacy of long-term Beta-blocker therapy for secondary prevention of long-term outcomes after coronary artery bypass grafting surgery. Circulation 2015; 131:2194-201.
- Philip F, Blackstone E, Kapadia SR. Impact of statin and beta blocker therapy on mortality after coronary artery bypass grafting surgery. Cardiovasc Diagn Ther 2015; 5:8-16

S.4. Numerator Statement: Number of patients undergoing isolated CABG who were discharged on beta blockers

S.6. Denominator Statement: Patients aged 18 years and older undergoing isolated CABG

S.8. Denominator Exclusions: Cases are removed from the denominator if there was an in-hospital mortality or if discharge beta blocker was contraindicated.

De.1. Measure Type: Process

S.17. Data Source: Registry Data

S.20. Level of Analysis: Clinician: Group/Practice, Facility

IF Endorsement Maintenance – Original Endorsement Date: May 09, 2007 **Most Recent Endorsement Date:** Jan 25, 2017

IF this measure is included in a composite, NQF (National Quality Forum) Composite#/title:

IF this measure is paired/grouped, NQF#/title:

De.4. IF PAIRED/GROUPED, what is the reason this measure must be reported with other measures to appropriately interpret results? N/A

Preliminary Analysis: Maintenance of Endorsement Measure

To maintain NQF endorsement endorsed measures are evaluated periodically to ensure that the measures still meet the NQF endorsement criteria ("maintenance"). The emphasis for maintaining endorsement is focused on how effective the measure is for promoting improvements in quality. Endorsed measures should have some experience from the field to inform the evaluation. The emphasis for maintaining endorsement is noted for each criterion.

Criteria 1: Importance to Measure and Report

1a. Evidence

Maintenance measures – less emphasis on evidence unless there is new information or change in evidence since the prior evaluation.

1a. Evidence. The evidence requirement for a *structure, process or intermediate outcome* measure is that it is based on a systematic review (SR) and grading of the body of empirical evidence where the specific focus of the evidence matches what is being measured. For measures derived from patient report, evidence also should demonstrate that the target population values the measured process or structure and finds it meaningful.

The developer provides the following evidence for this measure:

- **Systematic Review of the evidence specific to this measure?** ☒ **Yes** ☐ **No**
- **Quality, Quantity and Consistency of evidence provided?** ☐ **Yes** ☒ **No**
- **Evidence graded?** ☒ **Yes** ☐ **No**

Summary of prior review in 2016

- In 2016 the developer included the 2011 ACCF/AHA Guideline for Coronary Artery Bypass Graft Surgery. The recommendation stated:
    - Beta blockers should be prescribed to all CABG patients without contraindications at the time of hospital discharge (Class I Recommendation, Level of Evidence: C).
- The developer's summary of peer-reviewed literature provided during the last maintenance review supports that the utilization of beta-blockers at discharge confers a strong risk reduction in mortality.
- During the previous review, the Committee agreed the evidence was appropriate and consistent for the use of beta-blockers following isolated CABG.

Changes to evidence from last review

☒ **The developer attests that there have been no changes in the evidence since the measure was last evaluated.**

☐ **The developer provided updated evidence for this measure:**

Questions for the Committee:

- *The developer attests the underlying evidence for the measure has not changed since the last NQF endorsement review. Does the Committee agree the evidence basis for the measure has not changed and there is no need for repeat discussion and vote on Evidence?*

Guidance from the Evidence Algorithm

Process measure based on systematic (Box 3) à QQC not provided (Box 4) à Strong recommendation (Box 6) à Moderate

Preliminary rating for evidence: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

1b. Gap in Care/Opportunity for Improvement and 1b. Disparities

Maintenance measures – increased emphasis on gap and variation

1b. Performance Gap. The performance gap requirements include demonstrating quality problems and opportunity for improvement.

- On previous review, the Committee had asked the developer to include the number of patients included in the measure to help inform discussion of the performance gap. The developer has included the number of operations in this submission. Measure results calculated using registry

data for January-December 2018 (1037 participants and 151,805 operations) and January-December 2019 (999 participants and 150,773 operations).

Year	Mean	STD	IQR	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
2018	0.98	0.034	0.019	0.66	0.95	0.98	0.99	0.99	1.00	1.00	1.00	1.00	1.00	1.00
2019	0.98	0.043	0.016	0.00	0.96	0.98	0.99	0.99	1.00	1.00	1.00	1.00	1.00	1.00

Disparities

- Each year in the table below represents January-December.

Measures	2016	2017	2018	2019
All	98.60%	98.64%	98.79%	98.95%
Patient Gender	*	*	*	*
Male	98.67%	98.67%	98.84%	98.99%
Female	98.39%	98.53%	98.65%	98.79%
Age Groups	*	*	*	*
Age<75	98.69%	98.70%	98.89%	99.00%
Age>=75	98.23%	98.36%	98.39%	98.74%
Race Groups	*	*	*	*
White	98.73%	98.70%	98.86%	98.97%
Black	98.72%	98.75%	98.89%	98.95%
Other	97.56%	98.06%	98.21%	98.76%
Insurance, Age >=65	*	*	*	*
Medicare + Medicaid	98.42%	98.15%	98.45%	98.67%
Medicare + Commercial without Medicaid	98.70%	98.75%	98.78%	98.85%
Medicare without Medicaid/Commercial	98.13%	98.28%	98.59%	98.89%
Insurance, Age<65	*	*	*	*
Medicare/Medicaid	98.62%	98.67%	98.64%	98.83%
Commercial/HMO	98.80%	98.86%	99.07%	99.17%
None/Self Paid	99.17%	98.79%	99.04%	99.03%
Other	98.79%	98.48%	99.12%	99.08%

*cell intentionally left blank

Questions for the Committee:

- Is there a gap in care that warrants continued endorsement as a national performance measure?
- If you do not feel there is a gap that warrants continued active endorsement status should the measure be considered for Inactive Endorsement with Reserve Status?

Preliminary rating for opportunity for improvement: ☐ High ☐ Moderate ☒ Low ☐ Insufficient

RATIONALE: With a mean of 98%, an IQR of 0.016, and a median of 100% this measure appears topped out. Performance is uniformly high across disparity subgroups as well.

Committee Pre-evaluation Comments:

Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

1a. Evidence to Support Measure Focus: For all measures (structure, process, outcome, patient-reported structure/process), empirical data are required. How does the evidence relate to the specific structure, process, or outcome being measured? Does it apply directly or is it tangential? How does the structure, process, or outcome relate to desired outcomes? For maintenance measures –are you aware of any new studies/information that changes the evidence base for this measure that has not been cited in the submission? For measures derived from a patient report: Measures derived from a patient report must demonstrate that the target population values the measured outcome, process, or structure.

the evidence in support of this measure is good and there has not been substantive change since 2016

New recent study that supports evidence. Etchill, Eric W., and Glenn JR Whitman. "The Long-term Cardiac Benefits of β-Blockers After Coronary Bypass: Questioned but Not Disproven." The Annals of thoracic surgery 111.1 (2021): 75-76.

This is a maintenance measure, and no changes to previous evidence presented is reported.

Clinical guidelines and evidence continue to support utilization of beta-blockers.

The evidence supporting the measure focus has not changed much since last endorsement and remains moderate

1b. Performance Gap: Was current performance data on the measure provided? How does it

does it demonstrate disparities in the care?

Current performance on this measure is excellent on a national basis and the measure could be considered topped out. That being said, this process measure is part of the hospital/surgeon composite quality score and inactivating/retiring this measure could lead to a decreased focus on this

Does the data need to be further divided since the measure addresses patients over age 18? Current data

This measure is remarkably close to max out. Performance has been excellent and probably very limited room for improvement. However, I believe it is still an important measure to continue, since discontinuing may lead to reversal in the progress.

Data shows remarkably high performance on this measure. Minimal variation across subgroups shows little disparity.

Performance is almost perfect on this measure for >90% of entities. Performance is uniformly high (>98%) in all subgroups.

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability: Specifications and Testing

2b. Validity: Testing; Exclusions; Risk-Adjustment; Meaningful Differences; Comparability; Missing Data

2c. For composite measures: empirical analysis support composite approach

Reliability

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented. For maintenance measures – no change in emphasis – specifications should be evaluated the same as with new measures.

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers. For maintenance measures – less emphasis if no new testing data provided.

Validity

2b2. Validity testing should demonstrate the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For maintenance measures – less emphasis if no new testing data provided.

2b2-2b6. Potential threats to validity should be assessed/addressed.

Composite measures only:

2d. Empirical analysis to support composite construction. Empirical analysis should demonstrate that the component measures add value to the composite and that the aggregation and weighting rules are consistent with the quality construct.

Complex measure evaluated by Scientific Methods Panel? ☐ **Yes** ☒ **No**

Evaluators: NQF Staff

Scientific Acceptability Review

Questions for the Committee regarding reliability:

- Do you have any concerns that the measure can be consistently implemented (i.e., are measure specifications adequate)?
- The staff is satisfied with the reliability testing for the measure. Does the Committee have any concerns regarding reliability?

Questions for the Committee regarding validity:

- The staff raised concerns regarding the validity testing for the measure. What are your thoughts regarding the validity of the measure (e.g., exclusions, risk-adjustment approach, etc.)?

Preliminary rating for reliability:	☐ **High**	☒ **Moderate**	☐ **Low**	☐ **Insufficient**
Preliminary rating for validity:	☐ **High**	☐ **Moderate**	☐ **Low**	☒ **Insufficient**

Committee Pre-evaluation Comments:

Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2c)

2a1. Reliability-Specifications: Which data elements, if any, are not clearly defined? Which codes with descriptors, if any, are not provided? Which steps, if any, in the logic or calculation algorithm or other specifications (e.g., risk/case-mix adjustment, survey/sampling instructions) are not clear? What concerns do you have about the likelihood that this measure can be consistently implemented?

no concerns

no concerns

It has good history of reliability since initially approved.

No concerns with reliability.

No concerns with the specifications

2a2. Reliability - Testing: Do you have any concerns about the reliability of the measure?

no

no

none

No concerns with reliability.

I would prefer to see the actual distribution of the signal to noise ratio rather than the estimates driven by sample size to achieve different levels of reliability. But more importantly, the SNR does not directly address classification stability. If the main classification is the 3-tiered system, then what is the probability that an entity changes category using a resampling or Bayesian approach. I think this measure probably meets current reliability standards but I those are being rethought.?

2b1. Validity -Testing: Do you have any concerns with the testing results?

no

no concerns

No. The measure appears straight forward, and exclusions are as expected.

With such high performance, wonder if this measure continues to identify differences in quality.

The known groups validity testing was tautological? Entities will score significantly below the mean had lower scores. Also, I disagree that stability over a long time period is a test of validity. In fact, I think that perfect stability would be a sign of a bad measure, because it means that no change is possible.

2b2-3. Other Threats to Validity (Exclusions, Risk Adjustment) 2b2. Exclusions: Are the exclusions consistent with the evidence? Are any patients or patient groups inappropriately excluded from the measure? 2b3. Risk Adjustment: If outcome (intermediate, health, or PRO-based) or resource use performance measure: Is there a conceptual relationship between potential social risk factor variables and the measure focus? How well do social risk factor variables that were available and analyzed align with the conceptual description provided? Are all of the risk-adjustment variables present at the start of care (if not, do you agree with the rationale provided)? Was the risk adjustment (case-mix adjustment) appropriately developed and tested? Do analyses indicate acceptable results? Is an appropriate risk-adjustment strategy included in the measure?

The long-term risk reduction is conferred upon those patients continuing beta blocker therapy over the long term. This measure assesses the prescribing of beta blockers at hospital discharge

It does not appear included

No issues here.

No issues.

NA

2b4-7. Threats to Validity (Statistically Significant Differences, Multiple Data Sources, Missing Data) 2b4. Meaningful Differences: How do analyses indicate this measure identifies meaningful differences about quality? 2b5. Comparability of performance scores: If multiple sets of specifications: Do analyses indicate they produce comparable results? 2b6. Missing data/no response: Does missing data constitute a threat to the validity of this measure?

no concerns

Do lack of social risk factors contribute to validity?

None

Unconvinced that measure continues to identify meaningful differences in quality.

Classifying low performers as being statistically different from the average is problematic. If the average is 98% and a large entity has statistically different 97% performance, is that really low performance? I know this is the standard STS (Society of Thoracic Surgeons) method, but shouldn't some clinical/practical criteria be used in addition to statistical significance, especially when the distribution is so compressed?

Criterion 3. Feasibility

Maintenance measures – no change in emphasis – implementation issues may be more prominent

3. **Feasibility** is the extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.
    - The developer states that required data elements are generated or collected and used by healthcare personnel during provision of care. They are then abstracted from a record by someone other than the person obtaining the original information (e.g., chart abstraction). Some of the elements are available in EHRs (Electronic Health Records) or from other electronic sources.
    - Per the developer, the data elements in the measure have been standard in the STS Adult Cardiac Surgery Database for at least six years and some of them have been part of the database for more than 20 years. The database has more than 1,100 participants. Local availability of data elements will vary from full EHR (Electronic Health Record) capability to no availability; however, all data elements are submitted to the STS database in electronic format following a standard set of data specifications.
    - STS Adult Cardiac Surgery Database participants (single or group of surgeons) pay annual participant fees of $3,500 if majority of surgeons in the group are STS members and $4,750 if the majority are not STS members. In addition, there is a fee of $150 per member and $350 per

non-member for surgeons listed on the database's Participation Agreement. There are no additional costs for data collection specific to the measure.

Questions for the Committee:

- Are the required data elements routinely generated and used during care delivery?
- Are the required data elements available in electronic form, e.g., EHR or other electronic sources?

Preliminary rating for feasibility: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee Pre-evaluation Comments:

Criteria 3: Feasibility

3. Feasibility: Which of the required data elements are not routinely generated and used during care delivery? Which of the required data elements are not available in electronic form (e.g., EHR or other electronic sources)? What are your concerns about how the data collection strategy can be put into operational use?

no concerns

none

Elements are routinely collected during patient care and most data entered into STS Adult Cardiac Surgery Database.

No issues or concerns with feasibility.

This measure has already been implemented, so it is feasible.

Criterion 4: Usability and Use

Maintenance measures – increased emphasis – much greater focus on measure use and usefulness, including both impact/improvement and unintended consequences

4a. Use (4a1. Accountability and Transparency; 4a2. Feedback on measure)

4a. Use evaluate the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

4a.1. Accountability and Transparency. Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

Current uses of the measure

Publicly reported? ☒ **Yes** ☐ **No**

Current use in an accountability program? ☒ **Yes** ☐ **No** ☐ **UNCLEAR**

OR

Planned use in an accountability program? ☐ **Yes** ☐ **No**

Accountability program details

- This measure is part of a publicly reported composite (Perioperative Medications domain) as part of the voluntary STS Public Reporting of the isolated CABG composite. About 49.8% of the STS Adult Cardiac Surgery Database's 1,030 participants are voluntarily enrolled in the public reporting program.

4a.2. Feedback on the measure by those being measured or others. Three criteria demonstrate feedback: 1) those being measured have been given performance results or data, as well as assistance with interpreting the measure results and data; 2) those being measured, and other users have been given an opportunity to provide feedback on the measure performance or implementation; 3) this feedback has been considered when changes are incorporated into the measure

Feedback on the measure by those being measured or others

1. All Adult Cardiac Surgery Database participants receive quarterly feedback reports providing a detailed analysis of the participant's performance including benchmarking. Dashboard-type reporting on STS.org has been provided for real-time, online data updates to STS surgeon members. Participants also have access to a guide to help interpret performance results.
2. The adult cardiac surgeons from across the U.S. who comprise the STS Adult Cardiac Surgery Task Force meet periodically to discuss the participant reports and to consider potential enhancements to the ACSD. This feedback was one of the drivers for the real-time dashboard-type reporting recently implemented.
3. The developer did not provide any examples of feedback being considered when changes are incorporated into the measure.

Additional Feedback:

***Questions for the Committee**:*

- How have the performance results been used to further the goal of high-quality, efficient healthcare?
- How has the measure been vetted in real-world settings by those being measured or others?

Preliminary rating for Use: ☒ **Pass** ☐ **No Pass**

4b. Usability (4a1. Improvement; 4a2. Benefits of measure)

4b. Usability evaluate the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

4b.1 Improvement. Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated.

Improvement results

- Previous submission shows a rate of 97.96% for the period October 2011 – September 2012. The developer includes the overall rates of 98.62%, 98.80%, and 98.94%, for calendar years 2017, 2018, and 2019 respectively). This demonstrates improvement over time.

4b2. Benefits vs. harms. Benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

Unexpected findings (positive or negative) during implementation

- None reported.

Potential harms

- Potential harms include gaming and risk aversion. The developer states they control for these through a careful audit process and a robust risk-adjustment methodology.

***Questions for the Committee**:*

- How can the performance results be used to further the goal of high-quality, efficient healthcare?

Preliminary rating for Usability and use: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee Pre-evaluation Comments:

Criteria 4: Usability and Use

4a1. Use - Accountability and Transparency: How is the measure being publicly reported? Are the performance results disclosed and available outside of the organizations or practices whose performance is measured? For maintenance measures - which accountability applications are the measure being used for? For new measures - if not in use at the time of initial endorsement, is a credible plan for implementation provided? 4a2. Use - Feedback on the measure: Have those being measured been given performance results or data, as well as assistance with interpreting the measure results and data? Have those being measured or other users been given an opportunity to provide feedback on the measure performance or implementation? Has this feedback has been considered when changes are incorporated into the measure?

measure is publicly reported, and feedback is given to participating sites to incorporate into quality improvement initiatives

yes, STS interface

Appropriate feedback provided to participants

No concerns with use; publicly reported as part of accountability program. Feedback provided and received for the measure, no clear indication of incorporation in measure development.

The measure is being used as a standalone measure and part of a composite

4b1. Usability – Improvement: How can the performance results be used to further the goal of high-quality, efficient healthcare? If not in use for performance improvement at the time of initial endorsement, is a credible rationale provided that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations? 4b2. Usability – Benefits vs. harms: Describe any actual unintended consequences and note how you think the benefits of the measure outweigh them.

STS data has shown an improvement in performance in this measure over time. The potential harms for risk aversion and gaming are minimal with this particular measure

no concern

No harm noted. Still improving year to year, although now above 98%

Results over time show improvement, but remarkably high performance begs question of if continued value in furthering goal of high-quality. No other usability issues.

The vast majority of entities have exceptionally high performance, so it is unclear how useable the measure is in terms of driving QI (Quality Indicators). The performance rates in the last three 12-month

periods were 98.62%, 98.80% and 98.94%. Even if this is viewed as meaningful improvement, at what point will the measure be considered topped out? I think it may have arrived.

Criterion 5: Related and Competing Measures

Related or competing measures

NQF #0114 Risk-Adjusted Postoperative Renal Failure
NQF #0115 Risk-Adjusted Surgical Re-exploration
NQF #0116 Anti-Platelet Medication at Discharge
NQF #0118 Anti-Lipid Treatment Discharge
NQF #0119 Risk-Adjusted Operative Mortality for CABG
NQF #0127 Preoperative Beta Blockade
NQF #0129 Risk-Adjusted Postoperative Prolonged Intubation (Ventilation)
NQF #0130 Risk-Adjusted Deep Sternal Wound Infection
NQF #0131 Risk-Adjusted Stroke/Cerebrovascular Accident
NQF #0134 Use of Internal Mammary Artery (IMA) in Coronary Artery Bypass Graft (CABG)
NQF #0696 STS CABG Composite

Harmonization

- The related measures identified are NQF-endorsed measures developed by or with STS. All these measures are either components of NQF #0696 or are the overall composite NQF #0696. The developer indicates that they are harmonized.

Committee Pre-evaluation Comments: Criterion 5:

Related and Competing Measures

4. Related and Competing: Are there any related and competing measures? If so, are any specifications that are not harmonized? Are there any additional steps needed for the measures to be harmonized?

related measure is preoperative beta blockade. No competing measures

no additional steps

The related measures identified are NQF-endorsed measures developed by or with STS. The developer indicates that they are harmonized.

Multiple related measures, no concerns with harmonization.

No concerns

Public and Member Comments

Comments and Member Support/Non-Support Submitted as of: 01/26/2021

- Comment by: Society of Thoracic Surgeons

  STS Response to Preliminary Analyses for Measures 0117, 0127, 0134: Definitions for low- and high-performance groups

  The preliminary analyses for these three process measures found that “It is unclear how low and high-performance groups were defined” for known-group validity testing. This is in reference to

the “low performance.” “mid performance,” and “high performance” categories to which we refer in sect. 2b1.3 in the testing forms. The definitions of these categories are as described in sect. 2b4.1:

“Since higher value indicates better performance, an STS participant is designated as having higher/lower than average performance for the measure if the 95% CI [confidence interval] lies entirely above/below the STS average. The remaining participants are labeled as not distinguishable from the STS average performance. For the simplicity of this report, we call the three groups high performance, low performance, and mid performance, respectively.”

The high-, low-, and mid-performance groups are thus comparable to the STS “star rating” categories (“higher-than-expected,” “lower-than-expected,” “as-expected”), although the star ratings are applied to STS composite (outcome) measures only, not to individual process measures.

STS Response to Preliminary Analyses for Measures 0117, 0127, 0134: “Insufficient” ratings for Validity

We are aware that the NQF validity evaluation algorithm calls for other analyses (sensitivity, specificity, positive predictive value, negative predictive value) in addition to percent agreement. We believe, however, that the validity of our measures at the data element level is adequately demonstrated by the results of the exceptional external audit process that the STS has conducted annually since 2006.

The STS audit of the Adult Cardiac Surgery Database (ACSD) is designed to evaluate the accuracy, consistency, and comprehensiveness of data collection, and ultimately validate the integrity of the data stored in the Database. Each year, 10% of active ACSD participant sites are randomly selected for audit. In order to evaluate the comprehensiveness of the Database, a list of all cases that are submitted to our analytics center (Duke Clinical Research Institute [DCRI]) from three randomly selected months are compared to the hospital logs of all cases that are performed that year. The data managers provide the auditors with documentation of all cases performed. Each site must demonstrate an effective process to assure that all eligible cases are submitted to the Database.

DCRI randomly selects 20 CABG-only and 10 isolated valve cases that are performed in the calendar year for audit at each site; 12 CABG-only and 8 isolated valve cases are re-abstracted at each site. An over-sample is provided to allow for the possibility that a medical record cannot be located by the site and is therefore unavailable for re-abstraction.

A specified group of data variables are evaluated each year, utilizing the current version of the STS Adult Cardiac Surgery Data Specifications; the number of variables increases every year. (For example, 82 variables were evaluated in 2015; 86 in 2017; 91 in 2019.) Agreement rates are calculated for each of the individual variables, each variable category and overall. The overall aggregate agreement rate for the most recent five audits is shown in the table below:

Audit Year	Total Cases	Total Mismatch	Overall Aggregate Agreement Rate
2019	203,840	14,313	92.98%
2018	222,500	10,346	95.35%
2017	144,920	5,010	96.54%
2016	144,368	5,494	96.19%

2015	5,409	96.17%

These results, and the rigorous audit process through which they are obtained, demonstrate the accuracy and completeness of the data in the STS ACSD. This conclusion is further supported by comments received from our external auditors in each year's final audit report. Two examples follow:

[2015] "There were 141,047 total variables abstracted and there were 135,638 variables that matched, resulting in an overall agreement rate of 96.17% (95.73% in 2014). This overall performance rate reflects a high level of accuracy in data collection and evidence that the data contained in the ACSD are valid."

Source: The Society of Thoracic Surgeons Adult Cardiac Surgery Database Audit – Telligen Final Report. Telligen, December 2015.

[2018] "The overall aggregate agreement rate was 95.4%, demonstrating that the data contained in the ACSD is both comprehensive and highly accurate… The surgeons and staff that perform the data collection and submission to the ACSD were found to be committed to the STS goal of collecting quality data."

Source: The Society of Thoracic Surgeons Adult Cardiac Surgery Database Audit – Final Audit Report 2018. Cardiac Registry Support, LLC, November 2019.

In summary, we believe that the additional information provided here adequately demonstrates the validity of STS measures 0117, 0127, 0134 at the data element level, and will appreciate a reconsideration of the preliminary "insufficient" rating.

STS Response to Preliminary Analyses for Measures 0117 & 0134: "Low" ratings for Opportunity for Improvement

We understand but respectfully disagree with the assessment that these two STS measures are "topped out" and therefore subject to loss of endorsement. We ask that you please consider the following:

- The STS believes that these evidence based, guideline-directed measures are significantly responsible for the dramatic improvement we have demonstrated in outcomes and in process-of-care compliance, as documented in a 2019 Joint Commission Journal on Quality and Patient Safety article (1). Table 2 shows a 54% improvement in compliance with the Discharge Beta-Blocker measure (#0117) between 2002 and 2016, and a 32% improvement in compliance with the IMA Use measure (#0134) between 1998 and 2016.
- It is inappropriate to view these improvements as a rationale to remove endorsement for these measures and risk a deterioration in results due to the perception that these measures are no longer important. Cardiac surgeries are high-stakes procedures in which small errors or deviations from standardized care processes can lead to death. From our perspective, a residual 1-2 % failure rate for individual process measures is not acceptable.
- Cardiac surgery is comparable to the airline industry in that we must strive for high reliability; our goal is a 100% success rate.
- Even small failure rates may result in a participant rating below the STS average, providing the potential to identify statistically meaningful differences in performance.

- Furthermore, the continued use and endorsement of these measures does not contribute to an excessive data entry burden for clinicians or their staff. The data for these processes of care is routinely collected – in a data registry with over 95% participation in the U.S. – for the STS CABG Composite for which these are component measures, along with mortality and morbidity outcomes. Concerns related to measures becoming "topped out" are more relevant to non-registry measures for which data collection may require the allocation of additional resources.

We therefore believe that the "topped out" assessment for measures 0117 & 0134 is unwarranted and ask NQF staff and the Surgery Standing Committee to consider a higher Opportunity for Improvement rating for each measure.

1. Shahian DM. Professional Society Leadership in Health Care Quality: The Society of Thoracic Surgeons Experience. Joint Commission journal on quality and patient safety / Joint Commission Resources. 2019;45(7):466-79.

- No NQF have submitted support/non-support choices as of this date.

Scientific Acceptability Evaluation

Scientific Acceptability: Preliminary Analysis Form

Measure Number: 0117

Measure Title: Beta Blockade at Discharge

Type of measure:

☒ **Process** ☐ **Process: Appropriate Use** ☐ **Structure** ☐ **Efficiency** ☐ **Cost/Resource Use**
☐ **Outcome** ☐ **Outcome: PRO-PM (Patient Reported Outcomes Performance Measures)** ☐ **Outcome: Intermediate Clinical Outcome** ☐ **Composite**

Data Source:

☐ **Claims** ☐ **Electronic Health Data** ☐ **Electronic Health Records** ☐ **Management Data**
☐ **Assessment Data** ☐ **Paper Medical Records** ☐ **Instrument-Based Data** ☒ **Registry Data**
☐ **Enrollment Data** ☐ **Other**

Level of Analysis:

☒ **Clinician: Group/Practice** ☐ **Clinician: Individual** ☒ **Facility** ☐ **Health Plan**
☐ **Population: Community, County or City** ☐ **Population: Regional and State**
☐ **Integrated Delivery System** ☐ **Other**

Measure is:

☐ **New** ☒ **Previously endorsed (**NOTE: Empirical validity testing is expected at time of maintenance review; if not possible, justification is required.)

RELIABILITY: SPECIFICATIONS

1. **Are submitted specifications precise, unambiguous, and complete so that they can be consistently implemented? ☒ Yes ☐ No**

   Submission document: "MIF_xxxx" document, items S.1-S.22

   ***NOTE**: NQF staff will conduct a separate, more technical, check of eCQM specifications, value sets, logic, and feasibility, so no need to consider these in your evaluation.*

2. **Briefly summarize any concerns about the measure specifications.**

   No concerns.

RELIABILITY: TESTING

Submission document: "MIF_xxxx" document for specifications, testing attachment questions 1.1-1.4 and section 2a2

3. **Reliability testing level ☒ Measure score ☐ Data element ☐ Neither**
4. **Reliability testing was conducted with the data source and level of analysis indicated for this measure ☒ Yes ☐ No**
5. If score-level and/or data element reliability testing was NOT conducted or if the methods used were NOT appropriate, was **empirical VALIDITY testing** of patient-level data conducted?

   ☐ **Yes** ☐ **No**

6. **Assess the method(s) used for reliability testing**

   Submission document: Testing attachment, section 2a2.2

   - The measure's reliability was assessed appropriately, using a beta-binomial model of signal-to-noise ratio.

7. **Assess the results of reliability testing**

   - Reliability of the measure varies by number of eligible patients (denominator). 95% of the STS participants meet the 27-patient sample size necessary for 0.50 reliability and 76% meet the 62-patient sample size necessary for 0.70 reliability. The measure demonstrates at least moderate reliability for most providers.

   Submission document: Testing attachment, section 2a2.3

8. Was the method described and appropriate for assessing the proportion of variability due to real differences among measured entities? NOTE: If multiple methods used, at least one must be appropriate.

   Submission document: Testing attachment, section 2a2.2

   ☒ **Yes**

   ☐ **No**

   ☐ **Not applicable** (score-level testing was not performed)

9. Was the method described and appropriate for assessing the reliability of ALL critical data elements?

   Submission document: Testing attachment, section 2a2.2

   ☐ **Yes**

   ☐ **No**

   ☒ **Not applicable** (data element testing was not performed)

10. **OVERALL RATING OF RELIABILITY** (taking into account precision of specifications and all testing results):

    ☐ **High** (NOTE: Can be HIGH only if score-level testing has been conducted)

    ☒ **Moderate** (NOTE: Moderate is the highest eligible rating if score-level testing has not been conducted)

    ☐ **Low** (NOTE: Should rate LOW if you believe specifications are NOT precise, unambiguous, and complete or if testing methods/results are not adequate)

    ☐ **Insufficient** (NOTE: Should rate INSUFFICIENT if you believe you do not have the information you need to make a rating decision)

11. **Briefly explain rationale for the rating of OVERALL RATING OF RELIABILITY and any concerns you may have with the approach to demonstrating reliability.**

    Precise specifications (Box 1) → Empiric reliability testing (Box 2) → Testing at measure score level (Box 4) → Method described and appropriate (Box 5) → Level of confidence (Box 6) → Moderate

VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

12. **Please describe any concerns you have with measure exclusions.**

    Submission document: Testing attachment, section 2b2.

    No concerns.

13. **Please describe any concerns you have regarding the ability to identify meaningful differences in performance.**

    Submission document: Testing attachment, section 2b4.

    - The developer reports that for the period October 2014 – September 2014 around 80% of participants had performance indistinguishable from the STS average (95% CI), and the remaining participants performed differently.
        - 859 (82.9%) performed as expected
        - 94 (9.1%) had lower-than-expected performance
        - 83 (8%) had higher-than-expected performance
    - Given the uniformly high performance in the more recent data presented in the Performance Gap (1b) section, it would be expected that these meaningful differences would be more difficult to identify now.

14. **Please describe any concerns you have regarding comparability of results if multiple data sources or methods are specified.**
    Submission document: Testing attachment, section 2b5.
    - No concerns. There is only one data source/method for this measure.
15. **Please describe any concerns you have regarding missing data.**

    Submission document: Testing attachment, section 2b6.

    - No concerns.

16. **Risk Adjustment**

16a. **Risk-adjustment method** ☒ **None** ☐ **Statistical model** ☐ **Stratification**

16b. **If not risk-adjusted, is this supported by either a conceptual rationale or empirical analyses?**

☐ Yes ☐ No ☒ Not applicable (Process measure)

16c. **Social risk adjustment:**

16c.1 Are social risk factors included in risk model? ☐ Yes ☐ No ☒ Not applicable

16c.2 Conceptual rationale for social risk factors included? ☐ Yes ☒ No

16c.3 Is there a conceptual relationship between potential social risk factor variables and the measure focus? ☐ Yes ☒ No

16d. **Risk adjustment summary:**

16d.1 All of the risk-adjustment variables present at the start of care? ☐ Yes ☐ No

16d.2 If factors not present at the start of care, do you agree with the rationale provided for inclusion? ☐ Yes ☐ No

16d.3 Is the risk adjustment approach appropriately developed and assessed? ☐ Yes ☐ No

16d.4 Do analyses indicate acceptable results (e.g., acceptable discrimination and calibration) ☐ Yes ☐ No

16d.5. Appropriate risk-adjustment strategy included in the measure? ☐ Yes ☐ No

16e**. Assess the risk-adjustment approach**

N/A - No risk adjustment or risk stratification.

VALIDITY: TESTING

17. **Validity testing level:** ☐ **Measure score** ☐ **Data element** ☒ **Both**

18. **Method of establishing validity of the measure score:**

☐ **Face validity**

☒ **Empirical validity testing of the measure score**

☐ **N/A (score-level testing not conducted)**

19. **Assess the method(s) for establishing validity**

Submission document: Testing attachment, section 2b2.2

- Data element validity was assessed via the STS Adult Cardiac Surgery Database Audit, which randomly selected 10% of participating sites to evaluate the accuracy, consistency, and comprehensiveness of data collection. The audit process involves re-abstraction of data for 20 cases and comparison of 82 individual data elements with those submitted to the data warehouse. The results presented are from the 2015 audit. The method is appropriate for establishing data element validity.
- Measure score validity was examined using known-groups validity. For the measure score three performance groups were calculated and compared. The three groups had different proportions.
- Measure score validity was also examined using predictive validity/stability of measure score results over time. Data periods used were 10/2013 – 9/2014 and 10/2014 – 9/2015. Stability could be considered a test of reliability vs a test of validity of a measure. This methodology has been accepted to demonstrate validity in previous submissions.

20. **Assess the results(s) for establishing validity**

Submission document: Testing attachment, section 2b2.3

- The data element validity results provided demonstrate an overall agreement rate of 96.17% with most elements in the high 90% agreement range. Percent agreement alone does not provide enough information to fully evaluate data element validity (NQF validity algorithm, box 10).
- Known-group validity testing demonstrated that low-performance groups had lower observed rates and that high-performance groups had higher observed rates (91.1% vs 99.9%). It is unclear how low and high-performance groups were defined.
- Predicted validity/stability analysis demonstrated that among participants that were high performers during the first period, 76.1% were also high performance in the second period. 90% of mid-performers remained in the mid-performer category. Low performance showed more changes, with 49% remaining in the low-performer category in the second performance period.

21. **Was the method described and appropriate for assessing conceptually and theoretically sound hypothesized relationships?**

    Submission document: Testing attachment, section 2b1.

    ☒ **Yes**

    ☐ **No**

    ☐ **Not applicable** (score-level testing was not performed)

22. **Was the method described and appropriate for assessing the accuracy of ALL critical data elements?** *NOTE that data element validation from the literature is acceptable.*

    Submission document*: Testing attachment, section 2b1.*

    ☒ **Yes**

    ☐ **No**

    ☐ **Not applicable** (data element testing was not performed)

23. **OVERALL RATING OF VALIDITY taking into account the results and scope of all testing and analysis of potential threats.**

    ☐ **High** (NOTE: Can be HIGH only if score-level testing has been conducted)

    ☐ **Moderate** (NOTE: Moderate is the highest eligible rating if score-level testing has NOT been conducted)

    ☐ **Low** (NOTE: Should rate LOW if you believe that there <u>are</u> threats to validity and/or relevant threats to validity were not assessed OR if testing methods/results are not adequate)

    ☒ **Insufficient** (NOTE: For instrument-based measures and some composite measures, testing at both the score level and the data element level is required; if not conducted, should rate as INSUFFICIENT.)

24. **Briefly explain rationale for rating of OVERALL RATING OF VALIDITY and any concerns you may have with the developers' approach to demonstrating validity.**

    The information and testing provided is not sufficient to determine the validity of the composite measure. Would need additional statistics for the data element validity. Would need more information about the known-groups definition. Uncertain stability is an appropriate test for validity.

ADDITIONAL RECOMMENDATIONS

25. **If you have listed any concerns in this form, do you believe these concerns warrant further discussion by the multi-stakeholder Standing Committee? If so, please list those concerns**

Developer Submission

Brief Measure Information

NQF #: 0117

Corresponding Measures:

De.2. Measure Title: Beta Blockade at Discharge

Co.1.1. Measure Steward: The Society of Thoracic Surgeons

De.3. Brief Description of Measure: Percent of patients aged 18 years and older undergoing isolated CABG who were discharged on beta blockers

1b.1. Developer Rationale: The use of postoperative b-blockers is now known to protect patients both at one year and long term (greater than 5 years) from death following cardiac surgery. This effect is associated with a 46 % risk reduction in death at one –year and 35% risk reduction in mortality during long-term follow-up (see Chan below). The summary of peer reviewed literature cited below supports that the utilization of beta-blocker at discharge as conferring a strong risk reduction in mortality.

- Crystal E, Connolly SJ, Sleik K, et al. Interventions on prevention of postoperative atrial fibrillation in patients undergoing heart surgery: a meta-analysis. Circulation. 2002;106(1):75-80.
- 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.
- Maisel WH, Rawn JD, Stevenson WG. Atrial fibrillation after cardiac surgery. Ann Intern Med. 2001;135(12):1061-1073.
- 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.
- Welke KF, Ferguson TB, Coombs LP, et al. Validity of the Society of Thoracic Surgeons National Adult Cardiac Surgery Database. Ann Thorac Surg. 2004; 77:1137-1139.
- Charlson ME, Isom OW. Care after coronary-artery bypass surgery. N Engl J Med. 2003; 348:1456-63.
- Chen J, Radford MJ, Wang Y, Marciniak TA, Krumholz HM. Are beta-blockers effective in elderly patients who undergo coronary revascularization after acute myocardial infarction? Arch Intern Med. 2000; 160:947-52.
- Chan AYM, McAlister FA, Norris, CM, et al. Effect of B-Blocker use on outcomes after discharge in patients who underwent cardiac surgery. J Thorac Cardiovasc Surg. 2010; 140:182-7.
- Zhang H, Yuan X, Zhang H, et al. Efficacy of long-term Beta-blocker therapy for secondary prevention of long-term outcomes after coronary artery bypass grafting surgery. Circulation 2015; 131:2194-201.
- Philip F, Blackstone E, Kapadia SR. Impact of statin and beta blocker therapy on mortality after coronary artery bypass grafting surgery. Cardiovasc Diagn Ther 2015; 5:8-16

S.4. Numerator Statement: Number of patients undergoing isolated CABG who were discharged on beta blockers

S.6. Denominator Statement: Patients aged 18 years and older undergoing isolated CABG

S.8. Denominator Exclusions: Cases are removed from the denominator if there was an in-hospital mortality or if discharge beta blocker was contraindicated.

De.1. Measure Type: Process

S.17. Data Source: Registry Data

S.20. Level of Analysis: Clinician: Group/Practice, Facility

IF Endorsement Maintenance – Original Endorsement Date: May 09, 2007 **Most Recent Endorsement Date:** Jan 25, 2017

IF this measure is included in a composite, NQF Composite#/title:

IF this measure is paired/grouped, NQF#/title:

De.4. IF PAIRED/GROUPED, what is the reason this measure must be reported with other measures to appropriately interpret results? N/A

1. Evidence and Performance Gap – Importance to Measure and Report

Extent to which the specific measure focus is evidence-based, important to making significant gains in healthcare quality, and improving health outcomes for a specific high-priority (high-impact) aspect of healthcare where there is variation in or overall, less-than-optimal performance. ***Measures must be judged to meet all sub criteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form

117_NQF_evidence_attachment_BBDischarge_Fall_2020-637418380048053844.docx

1a.1 For Maintenance of Endorsement: Is there new evidence about the measure since the last update/submission?

Do not remove any existing information. If there have been any changes to evidence, the Committee will consider the new evidence. Please use the most current version of the evidence attachment (v7.1). Please use red font to indicate updated evidence.

No

1a. Evidence (sub criterion 1a)

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** 117

Measure Title: Beta Blockade at Discharge

IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here: 0696 STS CABG Composite Score

Date of Submission: **11/15/2020**

1a.1. This is a measure of: (*should be consistent with type of measure entered in De.1*)

Outcome

☐ Outcome:

☐ Patient-reported outcome (PRO):

PROs (Patient Reported Outcomes) include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors. (*A PRO-based performance measure is not a survey instrument. Data may be collected using a survey instrument to construct a PRO measure.)*

☐ Intermediate clinical outcome (*e.g., lab value*):

☒ Process:

- ☐ Appropriate use measure:

☐ Structure:

☐ Composite:

1a.2 LOGIC MODEL Diagram or briefly describe the steps between the healthcare structures and processes (e.g., interventions, or services) and the patient's health outcome(s). The relationships in the diagram should be easily understood by general, non-technical audiences. Indicate the structure, process or outcome being measured.

The summary of peer reviewed literature cited below supports that the utilization of beta-blocker at discharge as conferring a strong risk reduction in mortality. In addition, CABG is a frequently performed procedure, and a large number of patients undergo CABG yearly in the US. The development of post-operative atrial fibrillation consumes excess resources.

- Crystal E, Connolly SJ, Sleik K, et al. Interventions on prevention of postoperative atrial fibrillation in patients undergoing heart surgery: a meta-analysis. *Circulation.* 2002;106(1):75-80.
- 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.
- Maisel WH, Rawn JD, Stevenson WG. Atrial fibrillation after cardiac surgery. *Ann Intern Med.* 2001;135(12):1061-1073.
- 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.
- Welke KF, Ferguson TB, Coombs LP, et al. Validity of the Society of Thoracic Surgeons National Adult Cardiac Surgery Database. *Ann Thorac Surg.* 2004; 77:1137-1139.
- Charlson ME, Isom OW. Care after coronary-artery bypass surgery. *N Engl J Med*. 2003; 348:1456-63.
- Chen J, Radford MJ, Wang Y, Marciniak TA, Krumholz HM. Are beta-blockers effective in elderly patients who undergo coronary revascularization after acute myocardial infarction? *Arch Intern Med*. 2000; 160:947-52.
- Chan AYM, McAlister FA, Norris, CM, et al. Effect of B-Blocker use on outcomes after discharge in patients who underwent cardiac surgery*. J Thorac Cardiovasc Surg*. 2010; 140:182-7.
- Zhang H, Yuan X, Zhang H, et al. Efficacy of long-term Beta-blocker therapy for secondary prevention of long-term outcomes after coronary artery bypass grafting surgery. Circulation 2015; 131:2194-201.

- Philip F, Blackstone E, Kapadia SR. Impact of statin and beta blocker therapy on mortality after coronary artery bypass grafting surgery. Cardiovasc Diagn Ther 2015; 5:8-16.

1a.3 Value and Meaningfulness: IF this measure is derived from patient report, provide evidence that the target population values the measured ***outcome, process, or structure*** and finds it meaningful. (Describe how and from whom their input was obtained.)

****RESPOND TO ONLY ONE SECTION BELOW -EITHER 1a.2, 1a.3 or 1a.4) ****

1a.2 FOR OUTCOME MEASURES including PATIENT REPORTED OUTCOMES - Provide empirical data demonstrating the relationship between the outcome (or PRO) to at least one healthcare structure, process, intervention, or service.

1a.3. SYSTEMATIC REVIEW(SR) OF THE EVIDENCE (for INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURES, INCLUDING THOSE THAT ARE INSTRUMENT-BASED) If the evidence is not based on a systematic review go to section 1a.4) If you wish to include more than one systematic review, add additional tables.

What is the source of the systematic review of the body of evidence that supports the performance measure? A systematic review is a scientific investigation that focuses on a specific question and uses explicit, pre-specified scientific methods to identify, select, assess, and summarize the findings of similar but separate studies. It may include a quantitative synthesis (meta-analysis), depending on the available data. (IOM)

×Clinical Practice Guideline recommendation (with evidence review)

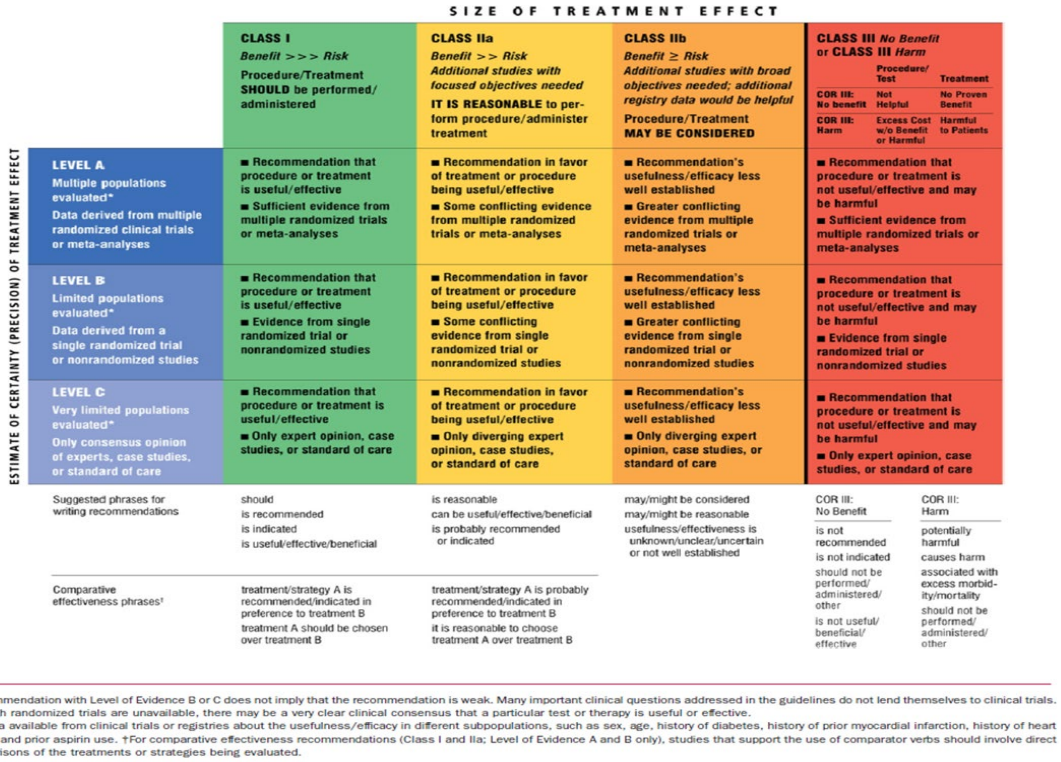
☐ US Preventive Services Task Force Recommendation

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*)

☐ Other

Systematic Review	Evidence
Source of Systematic Review: • Title • Author • Date	Hillis LD, Smith PK, Anderson JL, Bittl JA, Bridges CR, Byrne JG, et al. 2011 ACCF/AHA guideline for coronary artery bypass graft surgery. Circulation 2011;124: e652-735.

Systematic Review	Evidence
• Citation, including page number • URL	http://circ.ahajournals.org/content/124/23/e652
Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	Page e152 4.5. Perioperative Beta Blockers: Recommendations Class I Recommendation Beta blockers should be prescribed to all CABG patients without contraindications at the time of hospital discharge. (Level of Evidence: C)
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	Level C. Recommendation that procedure or treatment is useful/effective. Only expert opinion, case studies, or standard of care.
Provide all other grades and definitions from the evidence grading system	*
Grade assigned to the **recommendation** with definition of the grade	Class 1. See table below

Systematic Review	Evidence
Provide all other grades and definitions from the recommendation grading system	
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	Quantity – N/A Quality – see level of evidence above
Estimates of benefit and consistency across studies	*
What harms were identified?	*
Identify any recent studies conducted since the SR. Do the recent studies change the conclusions from the SR?	*

*cell intentionally left blank

1a.4 OTHER SOURCE OF EVIDENCE

If source of evidence is NOT from a clinical practice guideline, USPSTF, or systematic review, please describe the evidence on which you are basing the performance measure.

1a.4.1 Briefly SYNTHESIZE the evidence that supports the measure. A list of references without a summary is not acceptable.

1a.4.2 What process was used to identify the evidence?

1a.4.3. Provide the citation(s) for the evidence.

1b. Performance Gap

Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating:

- considerable variation, or overall, less-than-optimal performance, in the quality of care across providers; and/or
- Disparities in care across population groups.

1b.1. Briefly explain the rationale for this measure *(e.g., how the measure will improve the quality of care, the benefits or improvements in quality envisioned by use of this measure)*

If a COMPOSITE (e.g., combination of component measure scores, all-or-none, any-or-none), SKIP this question and answer the composite questions.

The use of postoperative b-blockers is now known to protect patients both at one year and long term (greater than 5 years) from death following cardiac surgery. This effect is associated with a 46 % risk reduction in death at one –year and 35% risk reduction in mortality during long-term follow-up (see Chan below). The summary of peer reviewed literature cited below supports that the utilization of beta-blocker at discharge as conferring a strong risk reduction in mortality.

- Crystal E, Connolly SJ, Sleik K, et al. Interventions on prevention of postoperative atrial fibrillation in patients undergoing heart surgery: a meta-analysis. Circulation. 2002;106(1):75-80.
- 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.
- Maisel WH, Rawn JD, Stevenson WG. Atrial fibrillation after cardiac surgery. Ann Intern Med. 2001;135(12):1061-1073.
- 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.
- Welke KF, Ferguson TB, Coombs LP, et al. Validity of the Society of Thoracic Surgeons National Adult Cardiac Surgery Database. Ann Thorac Surg. 2004; 77:1137-1139.
- Charlson ME, Isom OW. Care after coronary-artery bypass surgery. N Engl J Med. 2003; 348:1456-63.
- Chen J, Radford MJ, Wang Y, Marciniak TA, Krumholz HM. Are beta-blockers effective in elderly patients who undergo coronary revascularization after acute myocardial infarction? Arch Intern Med. 2000; 160:947-52.
- Chan AYM, McAlister FA, Norris, CM, et al. Effect of B-Blocker use on outcomes after discharge in patients who underwent cardiac surgery. J Thorac Cardiovasc Surg. 2010; 140:182-7.
- Zhang H, Yuan X, Zhang H, et al. Efficacy of long-term Beta-blocker therapy for secondary prevention of long-term outcomes after coronary artery bypass grafting surgery. Circulation 2015; 131:2194-201.

- Philip F, Blackstone E, Kapadia SR. Impact of statin and beta blocker therapy on mortality after coronary artery bypass grafting surgery. Cardiovasc Diagn Ther 2015; 5:8-16

1b.2. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis. *(This is required for maintenance of endorsement. Include mean, std dev, min, max, interquartile range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the sub-criterion on improvement (4b1) under Usability and Use.*

The measure was calculated using STS data for patients undergoing isolated CABG in two consecutive time periods January-December 2018 and January-December 2019. For each participant, the summary statistic provided is the proportion of eligible patients who receive beta blockers at discharge. An exact 95% exact binomial confidence interval was calculated for each participant´s observed proportion. A higher proportion indicates better performance. The percentiles were calculated after ordering the participants´ measures from the smallest to the largest. The 10th percentile value, for example, is the value that is larger than 10% of all participants.

Distribution of participant-specific observed proportions of eligible patients receiving the measure in January-December 2018 and January-December 2018.

Distribution	1/2018 - 12/2018	
Observed Proportion	1/2019 - 12/2019	
Observed Proportion		
# Participant	1037	999
# Operations	151805	150773
Mean	0.98	0.98
STD	0.034	0.043
IQR	0.019	0.016
0%	0.66	0.00
10%	0.95	0.96
20%	0.98	0.98
30%	0.99	0.99
40%	0.99	0.99
50%	1.00	1.00
60%	1.00	1.00
70%	1.00	1.00
80%	1.00	1.00
90%	1.00	1.00
100%	1.00	1.00

If the above table is not clearly displayed, please refer to the version included in the appendix for this measure.

1b.3. If no or limited performance data on the measure as specified is reported in 1b2, then provide a summary of data from the literature that indicates opportunity for improvement or overall, less than optimal performance on the specific focus of measurement.

N/A

1b.4. Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability. *(This is required for maintenance of endorsement. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included.) For measures that show high levels of performance, i.e., "topped out," disparities data may demonstrate an opportunity for improvement/gap in care for certain sub-populations. This information also will be used to address the sub-criterion on improvement (4b1) under Usability and Use.*

In the table below we provide trends over time of the measure at the patient level. Aggregate percentages of patients receiving the measure across four consecutive time periods are computed for relevant subgroups by age, gender, race, ethnicity, and insurance status.

	Jan 16-Dec 16	Jan 17-Dec 17	Jan 18-Dec 18	Jan 19-Dec 19
All	98.60%	98.64%	98.79%	98.95%
Patient Gender	98.67%	98.67%	98.84%	98.99%
Male				
Female	98.39%	98.53%	98.65%	98.79%
Age Groups	98.69%	98.70%	98.89%	99.00%
Age<75				
Age>=75	98.23%	98.36%	98.39%	98.74%
Race Groups	98.73%	98.70%	98.86%	98.97%
Race: White				
Race: Black	98.72%	98.75%	98.89%	98.95%
Race: Other	97.56%	98.06%	98.21%	98.76%
Insurance, Age>= 65				
Medicare+Medicaid	98.42%	98.15%	98.45%	98.67%
Medicare+Commercial without Medicaid	98.70%	98.75%	98.78%	98.85%
Medicare without Medicaid/Commercial	98.13%	98.28%	98.59%	98.89%
Insurance, Age<65				
Medicare/Medicaid	98.62%	98.67%	98.64%	98.83%
Commercial/HMO	98.80%	98.86%	99.07%	99.17%
None/Self Paid	99.17%	98.79%	99.04%	99.03%
Other	98.79%	98.48%	99.12%	99.08%

If the above table is not clearly displayed, please refer to the version included in the appendix for this measure.

1b.5. If no or limited data on disparities from the measure as specified is reported in 1b.4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations. Not necessary if performance data provided in 1b.4

N/A

2. Reliability and Validity—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. ***Measures must be judged to meet the sub criteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria.***

2a.1. Specifications The measure is well defined and precisely specified so it can be implemented consistently within and across organizations and allows for comparability. eMeasures should be specified in the Health Quality Measures Format (HQMF) and the Quality Data Model (QDM).

De.5. Subject/Topic Area *(check all the areas that apply):*

Cardiovascular, Surgery, Surgery: Cardiac Surgery

De.6. Non-Condition Specific *(check all the areas that apply):*

Safety, Safety: Medication

De.7. Target Population Category *(Check all the populations for which the measure is specified and tested if any):*

Adults, Elderly

S.1. Measure-specific Web Page *(Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials. Do not enter a URL linking to a home page or to general information.)*

https://www.sts.org/sites/default/files/STSAdultCVDataCollectionFormV4_20_2_GOLDEN006292020.pdf

S.2a. If this is an eMeasure, HQMF specifications must be attached. Attach the zipped output from the eMeasure authoring tool (MAT) - if the MAT was not used, contact staff. (Use the specification fields in this online form for the plain-language description of the specifications)

This is not an eMeasure Attachment**:**

S.2b. Data Dictionary, Code Table, or Value Sets *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*

No data dictionary Attachment**:**

S.2c. Is this an instrument-based measure (i.e., data collected via instruments, surveys, tools, questionnaires, scales, etc.)? Attach copy of instrument if available.

No, this is not an instrument-based measure Attachment**:**

S.2d. Is this an instrument-based measure (i.e., data collected via instruments, surveys, tools, questionnaires, scales, etc.)? Attach copy of instrument if available.

Not an instrument-based measure

S.3.1. For maintenance of endorsement: Are there changes to the specifications since the last updates/submission. If yes, update the specifications for S1-2 and S4-22 and explain reasons for the changes in S3.2.

No

S.3.2. For maintenance of endorsement, please briefly describe any significant changes to the measure specifications since last measure update and explain the reasons.

None

S.4. Numerator Statement *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome) DO NOT include the rationale for the measure.*

IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).

Number of patients undergoing isolated CABG who were discharged on beta blockers

S.5. Numerator Details *(All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*

IF an OUTCOME MEASURE, describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).

Number of isolated CABG procedures in which discharge beta blockers [DCBeta (STS Adult Cardiac Surgery Database Version 4.20)] is marked "yes"

S.6. Denominator Statement *(Brief, narrative description of the target population being measured)*

Patients aged 18 years and older undergoing isolated CABG

S.7. Denominator Details *(All information required to identify and calculate the target population/denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b.)*

IF an OUTCOME MEASURE, describe how the target population is identified. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).

Number of isolated CABG procedures excluding cases with an in-hospital mortality or cases for which discharge beta blocker use was contraindicated. The SQL code used to create the function used to identify cardiac procedures is provided in the Appendix.

S.8. Denominator Exclusions *(Brief narrative description of exclusions from the target population)*

Cases are removed from the denominator if there was an in-hospital mortality or if discharge beta blocker was contraindicated.

S.9. Denominator Exclusion Details *(All information required to identify and calculate exclusions from the denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b.)*

Mortality Discharge Status (DischMortStat), Mortality Date (MtDate), and Discharge Date (DischDt) indicate an in-hospital mortality; discharge beta blocker (DCBeta) marked as “Contraindicated”

S.10. Stratification Information *(Provide all information required to stratify the measure results, if necessary, including the stratification variables, definitions, specific data collection items/responses, code/value sets, and the risk-model covariates and coefficients for the clinically-adjusted version of the measure when appropriate – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format with at S.2b.)*

N/A

S.11. Risk Adjustment Type (Select type. Provide specifications for risk stratification in measure testing attachment)

No risk adjustment or risk stratification

If other:

S.12. Type of score:

Rate/proportion

If other:

S.13. 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

S.14. Calculation Algorithm/Measure Logic (*Diagram or describe the calculation of the measure score as an ordered sequence of steps including identifying the target population; exclusions; cases meeting the target process, condition, event, or outcome; time period for data, aggregating data; risk adjustment; etc.)*

Please refer to numerator and denominator sections for detailed information.

S.15. Sampling *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*

IF an instrument-based performance measure (e.g., PRO-PM), identify whether (and how) proxy responses are allowed.

N/A

S.16. Survey/Patient-reported data *(If measure is based on a survey or instrument, provide instructions for data collection and guidance on minimum response rate.)*

Specify calculation of response rates to be reported with performance measure results.

N/A

S.17. Data Source *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*

If other, please describe in S.18.

Registry Data

S.18. Data Source or Collection Instrument *(Identify the specific data source/data collection instrument (e.g., name of database, clinical registry, collection instrument, etc., and describe how data are collected.)*

IF instrument-based, identify the specific instrument(s) and standard methods, modes, and languages of administration.

STS Adult Cardiac Surgery Database Version 4.20

S.19. Data Source or Collection Instrument *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*

Available at measure-specific web page URL identified in S.1

S.20. Level of Analysis *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*

Clinician: Group/Practice, Facility

S.21. Care Setting *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*

Inpatient/Hospital

If other:

S.22. COMPOSITE Performance Measure - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*

N/A

2. Validity – See attached Measure Testing Submission Form

0117_NQF_testing_v7.1-BetaBlockade-11092020-637406040838713672-637418249507432033.docx

2.1 For maintenance of endorsement

Reliability testing: If testing of reliability of the measure score was not presented in prior submission(s), has reliability testing of the measure score been conducted? If yes, please provide results in the Testing attachment. Please use the most current version of the testing attachment (v7.1). Include information on all testing conducted (prior testing as well as any new testing); use red font to indicate updated testing.

No

2.2 For maintenance of endorsement

Has additional empirical validity testing of the measure score been conducted? If yes, please provide results in the Testing attachment. Please use the most current version of the testing attachment (v7.1). Include information on all testing conducted (prior testing as well as any new testing); use red font to indicate updated testing.

No

2.3 For maintenance of endorsement

Risk adjustment: For outcome, resource use, cost, and some process measures, risk-adjustment that includes social risk factors is not prohibited at present. Please update sections 1.8, 2a2, 2b1,2b4.3 and 2b5 in the Testing attachment and S.140 and S.11 in the online submission form. NOTE: These sections must be updated even if social risk factors are not included in the risk-adjustment strategy. You MUST use the most current version of the Testing Attachment (v7.1) -- older versions of the form will not have all required questions.

Yes - Updated information is included

Measure Testing (subcriteria 2a2, 2b1-2b6)

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b1-2b6)

Measure Number (*if previously endorsed*)**:** 0117

Measure Title: Beta Blockade at Discharge
Date of Submission: **8/1/2020**

Type of Measure:

Measure	Measure (continued)
☐ Outcome (*including PRO-PM*)	☐ Composite – *STOP – use composite testing form*
☐ Intermediate Clinical Outcome	☐ Cost/resource
☒ Process *(including Appropriate Use)*	☐ Efficiency
☐ Structure	*

*cell intentionally left blank

1. DATA/SAMPLE USED FOR ALL TESTING OF THIS MEASURE

Often the same data are used for all aspects of measure testing. In an effort to eliminate duplication, the first five questions apply to all measure testing. If there are differences by aspect of testing,(e.g., reliability vs. validity) be sure to indicate the specific differences in question 1.7.

1.1. What type of data was used for testing? (*Check all the sources of data identified in the measure specifications and data used for testing the measure. Testing must be provided for **all** the sources of data specified and intended for measure implementation. **If different data sources are used for the numerator and denominator, indicate N [numerator] or D [denominator] after the checkbox.***)

Measure Specified to Use Data From: (*must be consistent with data sources entered in S.17*)	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record
☐ claims	☐ claims
☒ registry	☒ registry
☐ abstracted from electronic health record	☐ abstracted from electronic health record
☐ eMeasure (HQMF (Health Quality Measures Format)) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☐ other:	☐ other:

1.2. If an existing dataset was used, identify the specific dataset (*the dataset used for testing must be consistent with the measure specifications for target population and healthcare entities being measured, e.g., Medicare Part A claims, Medicaid claims, other commercial insurance, nursing home MDS, home health OASIS, clinical registry*).

STS Adult Cardiac Surgery Database (ACSD) Version 4.2

1.3. What are the dates of the data used in testing? October 2014 – September 2015

1.4. What levels of analysis were tested? (*testing must be provided for **all** the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan*)

Measure Specified to Measure Performance of: (*must be consistent with levels entered in item S.20*)	**Measure Tested at Level of:**
☐ individual clinician	☐ individual clinician
☒ group/practice	☒ group/practice
☒ hospital/facility/agency	☒ hospital/facility/agency
☐ health plan	☐ health plan
☐ other:	☐ other:

1.5. How many and which measured entities were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of measured entities included in the analysis (e.g., size, location, type); if a sample was used, describe how entities were selected for inclusion in the sample*)

The calculation of the beta blockade at discharge measure of the 12 months from October 2014 to September 2015 used 139,564 operations from 1,036 STS ACSD participants.

Distribution of participant sample sizes (denominator), and observed proportion of patients receiving the measure (numerator/denominator)

Stat	N	% Beta Blockade at discharge
N	1036.0	1036.0
Mean	134.7	98.0
STD	107.2	3.9
IQR	115.0	2.4
0%	2.0	50.0
10%	37.0	94.3
20%	54.0	97.0
30%	71.0	98.1
40%	85.0	98.8
50%	103.5	99.3
60%	128.0	100.0
70%	156.0	100.0
80%	201.0	100.0
90%	268.5	100.0
100%	844.0	100.0

Distribution of participants by geographic regions

Region	# of participants
Midwest	296
Northeast	136
South	389
West	215

Our quality data are collected in the STS National Database at **participant**-level. As highlighted in the table below, over 92% of STS participants are **surgical practice groups that each have a "one to one" relationship with an individual hospital.** Therefore, with the exception of measures specifically identified as individual surgeon-focused (currently only the STS Individual Surgeon Composite for Adult Cardiac Surgery, NQF# 3030), STS performance measures are developed and validated at the STS participant level and do not require multiple levels of analysis.

Please note that the data in the table below includes all participants in the Adult Cardiac Surgery Database (ACSD) and is not specific to the subset of ACSD participants for whom data are reported for this specific measure.

Measures	**Distribution of STS "Participant" Contract Types in Adult Cardiac Surgery Database (11/2/2020)**	**Distribution Percentage of STS "Participant" Contract Types in Adult Cardiac Surgery Database (11/2/2020)**
Surgeon group only without hospital (including groups providing services at multiple hospitals), i.e., one-to-many	31	3.00%
Surgeon group w/individual hospital, i.e., one-to-one	952	92.40%
Surgeon group w/no hospital listed, i.e., new participant still being set up	2	0.20%

Measures	Distribution of STS "Participant" Contract Types in Adult Cardiac Surgery Database (11/2/2020)	Distribution Percentage of STS "Participant" Contract Types in Adult Cardiac Surgery Database (11/2/2020)
Individual surgeon	45	4.40%
Total US & Canada Participants	1030	100%

There is considerable sample size variation within and across different STS "participant" categories. To assure that our methodology is valid and reliable for any "participant" to whom we provide a score, we conduct sophisticated Markov Chain Monte Carlo (MCMC) simulations for all our measures to test their average reliability at different volume thresholds. STS estimates minimum sample size (i.e., case volume) thresholds, with their corresponding reliabilities, for each measure. We require that any participant receiving an STS score must have a volume of cases of the specific case type, during the prescribed analytic timeframe (i.e., typically 1 or 3 years), that assures an average reliability of 0.50, one of the highest measure reliability standards of which we are aware in all of healthcare.

1.6. How many and which patients were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis); if a sample was used, describe how patients were selected for inclusion in the sample*)

All eligible isolated operations were included except cases with an in-hospital mortality or cases for which discharge beta blocker use was contraindicated.

Measures	Effects	Overall N=139564
Age (years)	Median (IQR)	65.0 (58.0, 72.0)
*	Missing	0 (0.0%)
Sex	Male	105,326 (75.5%)
*	Female	34,176 (24.5%)
*	Missing	62 (0.0%)
Race - Asian	No	132,261 (94.8%)

Measures	Effects	Overall N=139564
*	Yes	4,294 (3.1%)
*	Missing	3,009 (2.2%)
Race - Black / African American	No	126,041 (90.3%)
*	Yes	10,517 (7.5%)
*	Missing	3,006 (2.2%)
Race - White	No	20,822 (14.9%)
*	Yes	115,801 (83.0%)
*	Missing	2,941 (2.1%)
Race - American Indian / Alaskan Native	No	135,676 (97.2%)
*	Yes	882 (0.6%)
*	Missing	3,006 (2.2%)
Race - Other	No	131,683 (94.4%)
*	Yes	4,495 (3.2%)
*	Missing	3,386 (2.4%)
Native Hawaiian / Pacific Islander	No	135,854 (97.3%)
*	Yes	641 (0.5%)
*	Missing	3,069 (2.2%)
Hispanic or Latino Ethnicity	No	122,462 (87.7%)
*	Yes	9,839 (7.0%)
*	Missing	7,263 (5.2%)
Insurance: Younger than 65	Medicare/Medicaid	17,491 (27.2%)
*	Commercial/HMO	38,339 (59.6%)
*	None/Self Paid	5,085 (7.9%)
*	Other	3,360 (5.2%)
Insurance: 65 or Older	Medicare+Medicaid	4,763 (6.3%)
*	Medicare+Commercial without Medicaid	41,526 (55.2%)
*	Medicare without Medicaid/Commercial	29,000 (38.5%)
Region	NORTHEAST	22,351 (16.0%)
*	SOUTH	60,956 (43.7%)
*	MIDWEST	34,154 (24.5%)
*	WEST	22,103 (15.8%)
Body Surface Area (m)	<1.5	1,766 (1.3%)
*	>=1.5 and <1.75	16,575 (11.9%)
*	>=1.75 and <2	47,745 (34.2%)

Measures	Effects	Overall N=139564
*	>=2	73,429 (52.6%)
*	Missing	49 (0.0%)
Diabetes	No Diabetes	72,185 (51.7%)
*	Diabetes - Noninsulin	41,308 (29.6%)
*	Diabetes - Insulin	24,619 (17.6%)
*	Diabetes - Other	369 (0.3%)
*	Diabetes - Missing Treatment	783 (0.6%)
*	Missing	300 (0.2%)
Hypertension	No	15,261 (10.9%)
*	Yes	124,016 (88.9%)
*	Missing	287 (0.2%)
Renal Function	Creatinine <1 mg/dL	67,662 (48.5%)
*	Creatinine 1-1.5 mg/dL	56,437 (40.4%)
*	Creatinine 1.5-2 mg/dL	8,187 (5.9%)
*	Creatinine 2-2.5 mg/dL	1,742 (1.2%)
*	Creatinine >2.5 mg/dL	1,317 (0.9%)
*	Dialysis	3,921 (2.8%)
*	Missing	298 (0.2%)
Dyslipidemia	No	16,601 (11.9%)
*	Yes	122,356 (87.7%)
*	Missing	607 (0.4%)
Chronic Lung Disease (CLD)	None	100,751 (72.2%)
*	Mild	14,875 (10.7%)
*	Moderate	6,713 (4.8%)
*	Severe	5,735 (4.1%)
*	5	6,864 (4.9%)
*	Missing	4,626 (3.3%)
Peripheral Vascular Disease (PVD)	No	119,135 (85.4%)
*	Yes	19,529 (14.0%)
*	Missing	900 (0.6%)
Cerebrovascular Disease (CVD)	No CVD	111,622 (80.0%)
*	CVD-NO CVA	27,942 (20.0%)
Endocarditis	No Endocarditis	139,331 (99.8%)
*	Treated Endocarditis	61 (0.0%)
*	Active Endocarditis	8 (0.0%)

Measures	Effects	Overall N=139564
*	Endocarditis - Missing Type	7 (0.0%)
*	Missing	157 (0.1%)
Acuity Status	Elective	52,969 (38.0%)
*	Urgent	80,674 (57.8%)
*	Emergent	5,745 (4.1%)
*	Emergent Salvage	156 (0.1%)
*	Missing	20 (0.0%)
Myocardial Infarction	No Prior MI (Myocardial Infarction)	65,332 (46.8%)
*	MI >21 days	26,411 (18.9%)
*	MI 8-21 days	6,673 (4.8%)
*	MI 1-7 days	34,686 (24.9%)
*	MI 6-24 hrs	3,285 (2.4%)
*	MI <= 6 hrs	1,679 (1.2%)
*	MI - Missing Timing	351 (0.3%)
*	Missing	1,147 (0.8%)
Cardiogenic Shock	No	137,887 (98.8%)
*	Yes	1,627 (1.2%)
*	Missing	50 (0.0%)
Preop IABP	No	129,589 (92.9%)
*	Yes	9,801 (7.0%)
*	Missing	174 (0.1%)
Congestive Heart Failure	No CHF (Congestive Heart Failure)	111,996 (80.2%)
*	CHF NYHA-I	2,314 (1.7%)
*	CHF NYHA-II	8,025 (5.8%)
*	CHF NYHA-III	9,566 (6.9%)
*	CHF NYHA-IV	5,513 (4.0%)
*	CHF Missing NYHA	926 (0.7%)
*	Missing	1,224 (0.9%)
Number of Diseased Coronary Vessels	None	130 (0.1%)
*	One	5,844 (4.2%)
*	Two	27,143 (19.4%)
*	Three	105,488 (75.6%)
*	Missing	959 (0.7%)
Left Main Disease > 50%	No	46,995 (33.7%)

Measures	Effects	Overall N=139564
*	Yes	44,312 (31.8%)
*	Missing	48,257 (34.6%)
Ejection Fraction (%)	Median (IQR)	55.0 (45.0, 60.0)
*	Missing	4,263 (3.1%)
Aortic Stenosis	No	132,712 (95.1%)
*	Yes	4,129 (3.0%)
*	Missing	2,723 (2.0%)
Mitral Stenosis	No	136,109 (97.5%)
*	Yes	687 (0.5%)
*	Missing	2,768 (2.0%)
Tricuspid Stenosis	No	136,229 (97.6%)
*	Yes	89 (0.1%)
*	Missing	3,246 (2.3%)
Pulmonic Stenosis	No	134,987 (96.7%)
*	Yes	29 (0.0%)
*	Missing	4,548 (3.3%)
Aortic Insufficiency	None	90,164 (64.6%)
*	Trivial	13,336 (9.6%)
*	Mild	10,506 (7.5%)
*	Moderate	2,061 (1.5%)
*	Severe	87 (0.1%)
*	N/A or Not Documented	22,255 (15.9%)
*	Missing	1,155 (0.8%)
Mitral Insufficiency	None	43,978 (31.5%)
*	Trivial	33,973 (24.3%)
*	Mild	32,654 (23.4%)
*	Moderate	8,504 (6.1%)
*	Severe	601 (0.4%)
*	N/A or Not Documented	18,915 (13.6%)
*	Missing	939 (0.7%)
Tricuspid Insufficiency	None	46,358 (33.2%)
*	Trivial	39,643 (28.4%)
*	Mild	25,503 (18.3%)
*	Moderate	3,954 (2.8%)
*	Severe	321 (0.2%)
*	N/A or Not Documented	22,536 (16.1%)

Measures	Effects	Overall N=139564
*	Missing	1,249 (0.9%)
Pulmonic Insufficiency	None	72,972 (52.3%)
*	Trivial	20,655 (14.8%)
*	Mild	6,531 (4.7%)
*	Moderate	539 (0.4%)
*	Severe	45 (0.0%)
*	N/A or Not Documented	37,264 (26.7%)
*	Missing	1,558 (1.1%)

*cell intentionally left blank

1.7. If there are differences in the data or sample used for different aspects of testing (e.g., reliability, validity, exclusions, risk adjustment), identify how the data or sample are different for each aspect of testing reported below.

We used the same dataset of isolated CABG operations from October 2014 to September 2015 for the entire report. The three exceptions are:

1. For validity testing and the comparison of participants over time, we used STS participants with procedures during both October 2013 - September 2014 and October 2014 - September 2015 time periods.
2. For the analysis of population disparities, current and over time, we used eligible patients from STS participants with procedures between October 2011 and September 2015 and defined relevant subgroups by age, gender, race, ethnicity, and insurance status.
3. For the analysis on the impact of exclusions, we included the cases with contraindication for beta blockade at discharge.

1.8 What were the social risk factors that were available and analyzed? For example, patient-reported data (e.g., income, education, language), proxy variables when social risk data are not collected from each patient (e.g., census tract), or patient community characteristics (e.g., percent vacant housing, crime rate) which do not have to be a proxy for patient-level data.

N/A (process measure, no risk model)

2a2. RELIABILITY TESTING

***Note**: If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – in 2a2.1 check critical data elements; in 2a2.2 enter "see section 2b2 for validity testing of data elements"; and skip 2a2.3 and 2a2.4.*

2a2.1. What level of reliability testing was conducted? (*may be one or both levels*)
☐ **Critical data elements used in the measure** (*e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements*)

☒ **Performance measure score** (e.g., *signal-to-noise analysis*)

2a2.2. For each level checked above, describe the method of reliability testing and what it tests (*describe the steps—do not just name a method; what type of error does it test; what statistical analysis was used*)

Reliability is conventionally defined as the proportion of variation in a measure that is due to true between-unit differences (i.e., signal) as opposed to random statistical fluctuations (i.e., noise). Equivalently, it is the squared correlation between a measurement and the true value. For this NQF submission, the measurement of interest is each participant's observed proportion. The true value is the proportion that would be observed hypothetically if the sample size was exceptionally large (i.e., infinite).

For the j-th participant, let n_j denote the number of eligible patients, let $y_j$denote the number of patients receiving beta-blockers, and let $\bar{y}_j = y_j/n_j$ denote the proportion of patients receiving beta-blockers. In addition, let μ_j denote the underlying true value of $\bar{y}_j$ To estimate reliability, we assumed the following hierarchical model for the data. At the first stage of the hierarchy, we assume that y_j is distributed according to a binomial distribution with sample size n_j and probability parameter μ_j. At the second stage of the hierarchy, we assumed that μ_j varies across participants according to a Beta distribution with mean $E[\mu_j] = \alpha/(\alpha+\beta)$ and var$[\mu_j] = \alpha\beta/[(\alpha+\beta)^2(\alpha+\beta+1)]$, where α and β are unknown parameters to be estimated from the data. The unknown parameters α and β were estimated via maximum likelihood using the BETABIN macro for SAS (Statistical Analysis System) software (BETABIN, version 2.2, 2005. Qi Statistics). The sample for this analysis included all **1,036 participants** and **139,564 eligible patients** in the main study period October 2014-September 2015. After estimating α and β, we then calculated the reliability that would be achieved if the measure were to be calculated on a sample size of 30 patients per participant. This estimated reliability was calculated as

$$\text{reliability} = [\text{corr}(\bar{y}, \mu)]^2 = \frac{1}{1 + (\hat{\alpha} + \hat{\beta})/n}$$

where $\hat{\alpha}$ and $\hat{\beta}$ denote maximum likelihood estimates of α and β, respectively, and n =30. Because reliability increases with n, and because the vast majority of STS participants have >30 eligible patients per year, the reliability calculated with n =30 patients per participant provides a conservative lower bound for the actual reliability that will be achieved when the measure is applied to STS data from a 1-year period. Using the above formula, we also calculated the sample size n required per participant to achieve reliability of at least 0.50, 0.60, and 0.70, and the proportion of STS participants with at least this number of eligible patients in the most recent 1-year testing sample.

2a2.3. For each level of testing checked above, what were the statistical results from reliability testing? (e.*g., percent agreement and kappa for the critical data elements; distribution of reliability statistics from a signal-to-noise analysis*)

Estimated parameter values of the beta distribution were $\hat{\alpha}$=26.162 and $\hat{\beta}$= 0.5024. The estimated reliability with 30 eligible patients per participant was $1/(1 + (26.162 + 0.5024)/30) = 0.53$.

Based on these estimated parameter values, a sample size of 27 eligible patients per participant is needed to attain reliability of 0.50 and a sample size of 62 eligible patients per participant is needed to attain reliability of 0.70. During October 2014-September 2015, 95% of STS participants met the minimum required sample size for 0.50 reliability and 76% of STS participants met the minimum required sample size for 0.70 reliability.

Measures	Reliability 0.50	Reliability 0.60	Reliability 0.70
Minimum required sample size per participant	27	40	62
Percent of participants meeting minimum sample size	95%	89%	76%

2a2.4 What is your interpretation of the results in terms of demonstrating reliability? (i*.e., what do the results mean and what are the norms for the test conducted?*)

Reliability is comparable to or better than other NQF-endorsed STS outcome measures. The proposed measure has adequate statistical reliability to be used for confidential feedback reporting as well as public reporting.

2b1. VALIDITY TESTING

2b1.1. What level of validity testing was conducted? (*may be one or both levels*)

☒ **Critical data elements** (*data element validity must address ALL critical data elements*)

☒ **Performance measure score**

☒ **Empirical validity testing**

☒ **Systematic assessment of face validity of performance measure score as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*) NOTE: Empirical validity testing is expected at time of maintenance review; if not possible, justification is required.

2b1.2. For each level of testing checked above, describe the method of validity testing and what it tests (*describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used)*

Critical data elements

Participating sites are randomly selected for participation in STS Adult Cardiac Surgery Database Audit, which is designed to evaluate the accuracy, consistency, and comprehensiveness of data collection and ultimately validate the integrity of the data contained in the database. Telligen has conducted audits on behalf of STS since 2006. In 2015, 10% of STS Adult Cardiac Surgery Database participants (N=107) were audited. The audit process involves re-abstraction of data for 20 cases and comparison of 82 individual data elements with those submitted to the data warehouse. Agreement rates are calculated for each of the 82 variables, each variable category and overall. In 2015 the overall aggregate agreement rate was

96.17%, demonstrating that the data contained in the STS Adult Cardiac Surgery Database are both comprehensive and highly accurate.

Performance measure score

We calculated and compared the observed proportions of patients receiving the measure in the three performance groups. The measure has good face value if the three groups have different proportions as expected.

Face validity also implies that the measure is regarded as useful and valid by its intended users, including providers, consumers, payers, and regulators. The measure was developed with a panel of surgeon experts and statisticians. We have had near-universal acceptance of this measure by all stakeholders, with few if any relevant suggestions for change.

In addition, we tested the predictive validity of the measure. Predictive validity means that the results of this measure are predictive of future performance. We assessed the extent to which performance on this STS measure remains stable over time. In other words, does the measure at one point in time accurately predict performance at some later time?

The tests on validity used the concept of performance outliers to be more formally introduced in 2b5: Participants were labeled as "low performance" if the 95% exact binomial confidence interval of its event rate lies entirely below the population average (in other words, the upper bound of the 95% CI < population average). Participants were labeled as "high performance" if the 95% confidence interval lies entirely above 1. The remaining participants were labeled mid performance.

For each of the performance groups from the earlier period, we calculated the group specific measure proportions in the later period.

2b1.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

Critical data elements

Database validity was evaluated by re-abstraction of defined variables from the medical records and comparison to submitted data. Agreement rates were calculated at the individual variable level, category level and overall. In the abridged table of 2015 Adult Cardiac Surgery Database (ACSD) audit results below, column one (CATEGORY) identifies the category each variable is assigned in the data specifications. The second column (FIELD NAME) represents the variable name and contains all the individual variables evaluated in the audit. The numerator column (NUM) represents the number of matches between the abstractors' findings and the responses submitted. The denominator column (DEN) is the total number of times the variable was abstracted, and the last column (Agreement Rate) contains the percentage agreement rates.

The overall agreement (data accuracy) rate for the 2015 ACSD audit was 96.17%.

Critical data elements and agreement rates relevant to this measure (Beta Blockade at Discharge) are shown in ***bold italics*** in the table below.

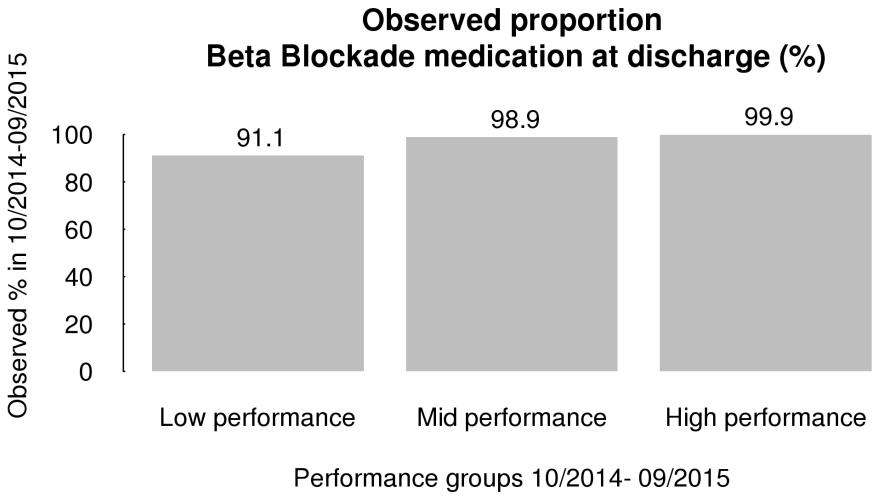
ACSD Aggregate Agreement Rates by Category, Field Name, and Overall (abridged)

CATEGORY	FIELD NAME	NUM	DEN	Agreement Rate
DEMOGRAPHICS	OVERALL ALL FIELDS	19094	19260	99.14%
DEMOGRAPHICS	Age (Age)	2129	2140	99.49%
DEMOGRAPHICS	Gender (Gender)	2131	2140	99.58%
DEMOGRAPHICS	White (RaceCaucasian)	2092	2140	97.76%
DEMOGRAPHICS	Black /African American (RaceBlack)	2129	2140	99.49%
DEMOGRAPHICS	Asian (RaceAsian)	2127	2140	99.39%
DEMOGRAPHICS	American Indian/Alaskan Native (RaceNativeAm)	2136	2140	99.81%
DEMOGRAPHICS	Native Hawaiian/Pacific Islander (RacNativePacific)	2137	2140	99.86%
DEMOGRAPHICS	Race (RaceOther)	2105	2140	98.36%
DEMOGRAPHICS	Hispanic or Latino or Spanish Ethnicity (Ethnicity)	2108	2140	98.50%
HOSPITALIZATION	OVERALL ALL FIELDS	6363	6420	99.11%
HOSPITALIZATION	Date of Admission (AdmitDt)	2118	2140	98.97%
HOSPITALIZATION	Date of Surgery (SurgDt)	2127	2140	99.39%
HOSPITALIZATION	Date of Discharge (DischDt)	2118	2140	98.97%
PRE-OPERATIVE MEDICATIONS	OVERALL ALL FIELDS	4086	4280	95.47%
PRE-OPERATIVE MEDICATIONS	Beta Blockers (MedBeta)	1982	2140	92.62%
PRE-OPERATIVE MEDICATIONS	Inotropes (MedInotr)	2104	2140	98.32%
OPERATIVE	OVERALL ALL FIELDS	4079	4280	95.30%
OPERATIVE	Status (Status)	2048	2140	95.70%
OPERATIVE	Appropriate Antibiotic Discontinuation (AbxDisc)	2031	2140	94.91%
CORONARY BYPASS	OVERALL ALL FIELDS	1408	1417	99.36%
CORONARY BYPASS	IMA Used for Grafts (IMAArtUs)	1306	1311	99.62%
CORONARY BYPASS	Reason for No IMA (NoIMARsn)	102	106	96.23%
POSTOPERATIVE	OVERALL ALL FIELDS	3968	4355	91.11%
POSTOPERATIVE	Postoperative Creatinine Level	1791	2137	83.81%
POSTOPERATIVE	Re-intubated During Hospital Stay	2107	2139	98.50%
POSTOPERATIVE	Additional Hours Ventilated	70	79	88.61%
POSTOPERATIVE EVENTS	OVERALL ALL FIELDS	16966	17010	99.74%
POSTOPERATIVE EVENTS	ReOp for Bleeding/Tamponade	2129	2130	99.95%
POSTOPERATIVE EVENTS	ReOp for Valvular Dysfunction	2130	2130	100.0%
POSTOPERATIVE EVENTS	ReOp for Graft Occlusion	2130	2130	100.0%

CATEGORY	FIELD NAME	NUM	DEN	Agreement Rate
POSTOPERATIVE EVENTS	ReOp for Other Cardiac	2122	2130	99.62%
POSTOPERATIVE EVENTS	ReOp for Other Non-Cardiac	2119	2130	99.48%
POSTOPERATIVE EVENTS	Deep Sternal Infection	2098	2100	99.90%
POSTOPERATIVE EVENTS	Postoperative Stroke > 24 Hours	2127	2130	99.86%
POSTOPERATIVE EVENTS	Renal Failure	2111	2130	99.11%
MORTALITY	OVERALL ALL FIELDS	6480	6572	98.60%
MORTALITY	Mortality (Mortalty)	2116	2140	98.88%
MORTALITY	Discharge Status (MtDCStat)	2138	2140	99.91%
MORTALITY	Status at 30 Day After Surgery (Mt30Stat)	2079	2140	97.15%
MORTALITY	Operative Death (MtOpD)	147	152	96.71%
DISCHARGE	OVERALL ALL FIELDS	8189	8396	97.53%
DISCHARGE	ADP Inhibitors (DCADP)	2063	2099	98.28%
DISCHARGE	Aspirin (DCASA)	2047	2099	97.52%
DISCHARGE	***Beta Blockers (DCBeta)***	***2040***	***2099***	***97.19%***
DISCHARGE	Lipid Lowering (DCLipid)	2039	2099	97.14%
	OVERALL ALL FIELDS	**135638**	**141047**	**96.17%**

Performance measure score

STS participants deemed high performers by this measure have (on average) high rates of beta blockade at discharge. Thus, differences in performance were clinically meaningful as well as statistically significant. This is illustrated in the figure below using data from October 2014 to September 2015. Compared to participants who were deemed as having lower than average performance, those with better-than-average performance had higher rate of beta blockade at discharge (99.9% vs. 91.1%).

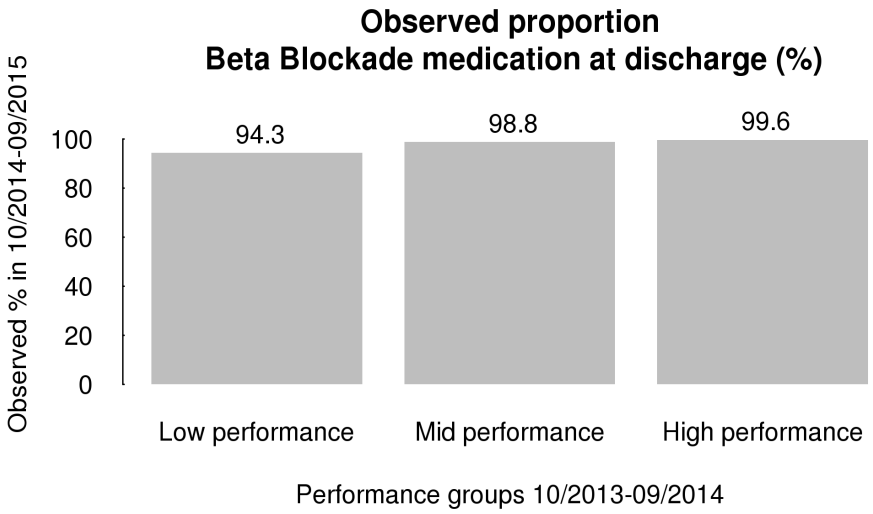


The predicted validity analysis was restricted to a sample of 1012 STS participants with patients receiving the measure in both time periods: October 2013 – September 2014 and October 2014 - September 2015. Among participants who were high performance centers in October 2013 – September 2014, 76.1% of them were also high performers for October 2014 - September 2015. For comparison, only 5.2% of participants who were mid performers in October 2013 – September 2014 became high performers in October 2014 - September 2015. Thus, participants who performed better than average in October 2013 – September 2014 were over 14 times more likely to be identified as better performers in the next year. Similarly, participants who were low performance entities in the early year were more likely to remain low performers in the later year. Two participants jumped from low to high performing status (or vice versa) between the two adjacent 12-month periods. Thus, a consumer may reasonably expect that a high or low performer will likely be the same or became average in the near future, and a mid-performer is likely to remain average.

Change in performance categories between two time periods

Measures	**10/2014 – 09/ 2015:** Low performance	**10/2014 – 09/ 2015:** Mid performance	**10/2014 – 09/ 2015:** High performance
10/2013 -09/2014: Low performance	50	50	2
10/2013 -09/2014: Mid performance	39	780	45
10/2013 -09/2014: High performance	0	11	35

For each of the performance groups in the earlier period, we also calculated its aggregated proportion of patients receiving the measure in the later period. The aggregated proportions in the later periods were 99.6%, 98.8%, and 94.3% for the high, mid, and low performance groups from the earlier period.



2b1.4. What is your interpretation of the results in terms of demonstrating validity? (i.*e., what do the results mean and what are the norms for the test conducted?*)

The high (96.17%) overall agreement rate for critical data elements in the STS Adult Cardiac Surgery Database reflects a high level of accuracy in data collection and evidence that the data contained in this database are valid.

The performance measure test results show that the measure reflects the proportion of patients who were discharged on beta blockers as designed, and that the past measure can be used to predict future performance. Together with face value, they support the validity of the measure.

_____

2b2. EXCLUSIONS ANALYSIS

NA ☐ no exclusions — *skip to section 2b4*

2b2.1. Describe the method of testing exclusions and what it tests (*describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used*)

We excluded from the analysis cases if there was an in-hospital mortality or if beta blocker was contraindicated. We believe this is a clinically appropriate exclusion and is necessary to make the measure a consistent performance measure for the comparison across participants. The exclusion is precisely defined and specified.

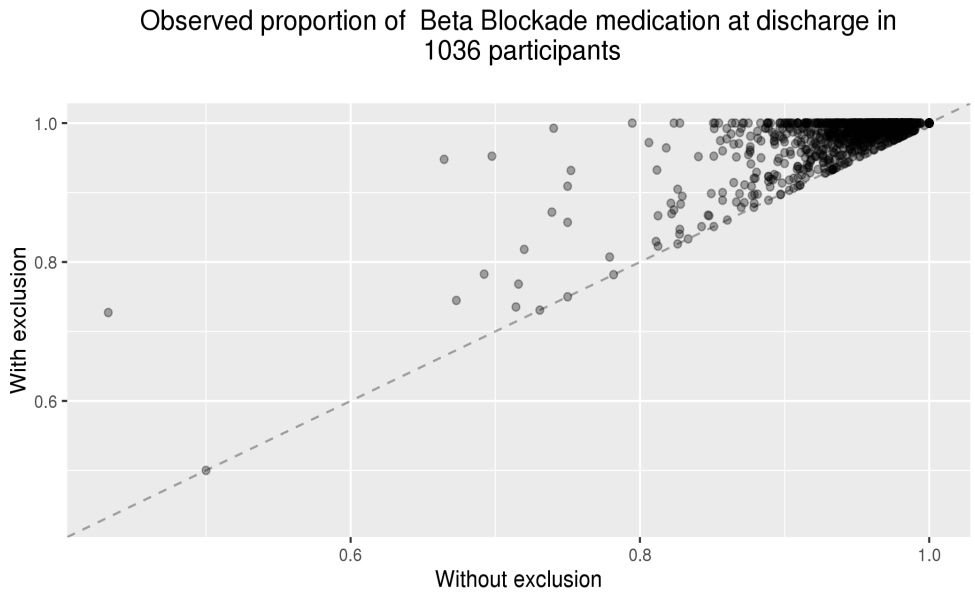
To show the impact of this exclusion, and how the measure would be distributed without it, we calculated and compared the distributions of the measure with and without the current exclusion criteria, with the exception of in-hospital deaths, that were excluded in all analyses.

2b2.2. What were the statistical results from testing exclusions? (*include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores*)

Distribution of participant-specific observed proportion of patients receiving the measure in October 2014 - September 2015 with and without the exclusion

Distribution	10/2014 – 09/2015 Observed proportion with exclusion	10/2014 – 09/2015 Observed proportion without exclusion
# Participant	1036	1036
# Operations	139564	144880
Mean	0.98	0.94
STD	0.039	0.051
IQR	0.024	0.046
0%	0.50	0.43
10%	0.94	0.89
20%	0.97	0.92
30%	0.98	0.94
40%	0.99	0.95
50%	0.99	0.96
60%	1.00	0.96
70%	1.00	0.97
80%	1.00	0.98
90%	1.00	0.99
100%	1.00	1.00
Midwest	296	296
Northeast	136	136
South	389	389
West	215	215
Low performance	94, 9.1%	103, 9.9%
Mid performance	859, 82.9%	830, 80.1%
High performance	83, 8.0%	103, 9.9%

Comparison of measure scores with and without the exclusion



The Spearman rank correlation of the measures with and without the exclusion is 0.54. The Pearson correlation is 0.75.

2b2.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results? (*i.e., the value outweighs the burden of increased data collection and analysis. Note: **If patient preference is an exclusion**, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion*)

For the measure to consistently quantify the quality per its definition, it is necessary to exclude cases if there was an in-hospital mortality or if discharge beta blocker was contraindicated. It has an impact on the results for many participants, and the results would be distorted without these appropriate exclusions.

2b3. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES
If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b5.

2b3.1. What method of controlling for differences in case mix is used?

☒ **No risk adjustment or stratification**

☐ **Statistical risk model with risk factors**

☐ **Stratification by risk categories**

☐ **Other,**

2b3.1.1 If using a statistical risk model, provide detailed risk model specifications, including the risk model method, risk factors, coefficients, equations, codes with descriptors, and definitions.

2b3.2. If an outcome or resource use component measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (case mix) is not

needed to achieve fair comparisons across measured entities.

2b3.3a. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or social risk factors) used in the statistical risk model or for stratification by risk (*e.g., potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10; correlation of x or higher; patient factors should be present at the start of care*) **Also discuss any "ordering" of risk factor inclusion**; for example, are social risk factors added after all clinical factors?

2b3.3b. How was the conceptual model of how social risk impacts this outcome developed? Please check all that apply:

- ☐ **Published literature**
- ☐ **Internal data analysis**
- ☐ **Other (please describe)**

2b3.4a. What were the statistical results of the analyses used to select risk factors?

2b3.4b. Describe the analyses and interpretation resulting in the decision to select social risk factors *(e.g., prevalence of the factor across measured entities, empirical association with the outcome, contribution of unique variation in the outcome, assessment of between-unit effects and within-unit effects.)* **Also describe the impact of adjusting for social risk (or not) on providers at high or low extremes of risk.**

2b3.5. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (*describe the steps—do not just name a method; what statistical analysis was used*)

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.
If stratified, skip to 2b3.9

2b3.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

2b3.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

2b3.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:

2b3.9. Results of Risk Stratification Analysis:

2b3.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)? (i.*e., what do the results mean and what are the norms for the test conducted*)

2b3.11. Optional Additional Testing for Risk Adjustment (*not required, but would provide additional support of adequacy of risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed*)

2b4. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

2b4.1. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified (*describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided related to performance gap in 1b)*

The summary statistic provided is the participant's observed proportion of eligible patients who receive beta blocker at discharge.

The degree of uncertainty surrounding an STS participant's beta blockade at discharge measure estimate is indicated by the 95% exact binomial confidence interval (CI) of its observed proportion. Point estimates and CI's of the observed proportion for an individual STS participant are reported along with a comparison to the STS average proportion of the study period. A performance category interpretation is also given to STS participants. **Since higher value indicates better performance**, an STS participant is designated as having higher/lower than average performance for the measure if the 95% CI lies entirely **above/below** the STS average. The remaining participants are labeled as not distinguishable from the STS average performance. For the simplicity of this report, we call the three groups 'high performance', 'low performance' and 'mid performance', respectively.

The method is equivalent to performing an exact binomial test with the null hypothesis that the participant has the same proportion of patients receiving the measure as the population average. Those with a test p-value smaller than 0.05 are the low and high-performance groups.

2b4.2. What were the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities? (e.g., *number and percentage of entities with scores that were statistically significantly different from mean or some benchmark, different from expected; how was meaningful difference defined*)
As shown in the table below, the proportion of STS ACSD participants performing better and worse than STS average has remained similar over the last two 12-month periods. On average, more than 80% of the participants have performance indistinguishable from the STS average, and the remaining participants have performed differently.

Distribution	10/2013 – 09/2014 Observed Proportion	10/2014 – 09/2015 Observed Proportion
# Participant	1058	1036
# Operations	139921	139564
Low performance	109, 10.3%	94, 9.1%
Mid performance	902, 85.3%	859, 82.9%
High performance	47, 4.4%	83, 8.0%

2b4.3. What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities? (i*.e., what do the results mean in terms of statistical and meaningful differences?*)

The statistical test and the construction of confidence interval are widely used and accepted. The participants identified as having performed differently from the average likely have true performance characteristics that are different. The identified differences in performance are both statistically significant and clinically meaningful. The surgeon panel and users are satisfied with the number of outliers the measure detects.

2b5. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS

If only one set of specifications, this section can be skipped.

Note*: This item is directed to measures that are risk-adjusted (with or without social risk factors)* ***OR*** *to measures with more than one set of specifications/instructions (e.g., one set of specifications for how to identify and compute the measure from medical record abstraction and a different set of specifications for claims or eMeasures). It does not apply to measures that use more than one source of data in one set of specifications/instructions (e.g., claims data to identify the denominator and medical record abstraction for the numerator).* ***Comparability is not required when comparing performance scores with and without social risk factors in the risk adjustment model. However, if comparability is not demonstrated for measures with more than one set of specifications/instructions, the different specifications (e.g., for medical records vs. claims) should be submitted as separate measures.***

2b5.1. Describe the method of testing conducted to compare performance scores for the same entities across the different data sources/specifications (*describe the steps—do not just name a method; what statistical analysis was used*)

2b5.2. What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications? (*e.g., correlation, rank order*)

2b5.3. What is your interpretation of the results in terms of the differences in performance measure scores for the same entities across the different data sources/specifications? (i.*e., what do the results mean and what are the norms for the test conducted*)

2b6. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b6.1. Describe the method of testing conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and non-responders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)
Due to great data quality, the source fields required by beta blockade at discharge had only 0.3% missing in the latest measure time window. We calculated the overall rate of missing as well as missing rates across all participants. In the implementation, missing data are imputed to "no". In addition, participants with greater than 5% missing data are excluded from the calculation of the measure.

2b6.2. What is the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data? (*e.g., results of sensitivity analysis of the effect of various rules for missing data/nonresponse; if no empirical sensitivity analysis, identify the approaches for handling missing data that were considered and pros and cons of each*)

Overall, 0.3% of data were missing. 99% of participants had missing rate of 4% or lower. Ten out of 1048 participants were not included because of having missing rates higher than 5%.

2b6.3. What is your interpretation of the results in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and non-responders) and how the specified handling of missing data minimizes bias**?** (i*.e., what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; if no empirical analysis, provide rationale for the selected approach for missing data*)

The rates of missing data in the STS Adult Cardiac Surgery Database were exceptionally low and are getting lower. We therefore concluded that systematic missing data did not lead to bias in our measure.

3. Feasibility

Extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

3a. Byproduct of Care Processes

For clinical measures, the required data elements are routinely generated and used during care delivery (e.g., blood pressure, lab test, diagnosis, medication order).

3a.1. Data Elements Generated as Byproduct of Care Processes.

Generated or collected by and used by healthcare personnel during the provision of care (e.g., blood pressure, lab value, diagnosis, depression score), Abstracted from a record by someone other than person obtaining original information (e.g., chart abstraction for quality measure or registry)

If other:

3b. Electronic Sources

The required data elements are available in electronic health records or other electronic sources. If the required data are not in electronic health records or existing electronic sources, a credible, near-term path to electronic collection is specified.

3b.1. To what extent are the specified data elements available electronically in defined fields (*i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields*) Update this field for **maintenance of endorsement**.

Some data elements are in defined fields in electronic sources

3b.2. If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources. For **maintenance of endorsement**, if this measure is not an eMeasure (eCQM), please describe any efforts to develop an eMeasure (eCQM).

The STS Adult Cardiac Surgery Database (ACSD) has 1,030 participants as of August 2020, and local availability of data elements in electronic format will vary across institutions. Some institutions may have full EHR capability while others may have partial, or no availability. However, all data elements from participating institutions are submitted to the STS ACSD in electronic format following a standard set of data specifications. The majority of participating institutions obtain data entry software products that are certified for the purposes of collecting STS ACSD data elements.

3b.3. If this is an eMeasure, provide a summary of the feasibility assessment in an attached file or make available at a measure-specific URL. Please also complete and attach the NQF Feasibility Score Card.

Attachment:

3c. Data Collection Strategy

Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, costs associated with fees/licensing of proprietary measures) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use). For eMeasures, a feasibility assessment addresses the data elements and measure logic and demonstrates the eMeasure can be implemented or feasibility concerns can be adequately addressed.

3c.1. Required for maintenance of endorsement. Describe difficulties (as a result of testing and/or operational use of the measure) regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.

IF instrument-based, consider implications for both individuals providing data (patients, service recipients, respondents) and those whose performance is being measured.

The data elements included in this measure have been standard in the STS Adult Cardiac Surgery Database for at least 6 years and some of them have been part of the database for more than 20 years. The variables are considered to be data elements that are readily available and already collected as part of the process of providing care.

3c.2. Describe any fees, licensing, or other requirements to use any aspect of the measure as specified *(e.g., value/code set, risk model, programming code, algorithm)*.

Data Collection:

There are no additional costs for data collection specific to this measure for those presently using and participating in the STS Adult Cardiac Surgery Database. Costs to develop and maintain the measure included volunteer cardiothoracic surgeon time, STS staff time, and Duke Clinical Research Institute statistician and project management time.

Other fees:

STS Adult Cardiac Surgery Database participants (generally a group of surgeons) pay annual participant fees of $3,500 or $4,750, depending on whether the majority of surgeons in a participant group are STS members. As a benefit of STS membership, the member-majority participants are charged the lesser of the two fees. Also, member-majority participants pay an additional fee of $150 per surgeon; non-member-majority participants pay an additional fee of $350 per surgeon.

4. Usability and Use

Extent to which potential audiences (e.g., consumers, purchasers, providers, policy makers) are using or could use performance results for both accountability and performance improvement to achieve the goal of high-quality, efficient healthcare for individuals or populations.

4a. Accountability and Transparency

Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

4.1. Current and Planned Use

NQF-endorsed measures are expected to be used in at least one accountability application within 3 years and publicly reported within 6 years of initial endorsement in addition to performance improvement.

Specific Plan for Use	Current Use (for current use provide URL)
*	Public Reporting STS Public Reporting https://www.sts.org/registries/sts-public-reporting STS Public Reporting https://www.sts.org/registries/sts-public-reporting Quality Improvement (Internal to the specific organization) STS Adult Cardiac Surgery Database https://www.sts.org/registries-research-center/sts-national-database/adult-cardiac-surgery-database

*cell intentionally left blank

4a1.1 For each CURRENT use, checked above (update for maintenance of endorsement), provide:

- Name of program and sponsor
- Purpose
- Geographic area and number and percentage of accountable entities and patients included
- Level of measurement and setting

Voluntary STS Public Reporting – approximately 79% of STS Adult Cardiac Surgery Database participants are enrolled as of October 2020.
This measure is publicly reported as a component of the Perioperative Medications domain of the isolated CABG composite.
(https://publicreporting.sts.org/acsd)
STS Adult Cardiac Surgery Database Participant Feedback Reports provide performance results for this measure to participants. (see details in 4a2.1.1 below)

4a1.2. If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing) what are the reasons? (*e.g., Do policies or actions of the developer/steward or accountable entities restrict access to performance results or impede implementation?*)
N/A

4a1.3. If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes -- any accountability application within 3 years and publicly reported within 6 years of initial endorsement. (*Credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.*)

N/A

4a2.1.1. Describe how performance results, data, and assistance with interpretation have been provided to those being measured or other users during development or implementation.

How many and which types of measured entities and/or others were included? If only a sample of measured entities were included, describe the full population and how the sample was selected.

As of November 2020, there are 1,030 active U.S. and Canadian participants in the STS Adult Cardiac Surgery Database (ACSD). A "participant" is generally a group of cardiothoracic surgeons who agree to submit case records for analysis and comparison with benchmarking data for quality improvement initiatives. At the option of the surgical group, the ACSD participant can include a hospital and/or associated anesthesiologists. It is for this reason that we have indicated (on the Specifications tab, question #S.20) that this measure is specified/tested for both the "clinician: group/practice" and "facility" levels of analysis.

(For more information on STS "participants," see our response to 1.5 in the measure testing form.)

All ACSD participants receive quarterly data reports with their performance results, reported in an easy-to-understand format. The participant's score is illustrated graphically in relation to the 25th, 50th and 75th percentiles of the distribution across all participants who were eligible for inclusion in that quarter´s analysis

and is also accompanied by the 95% Bayesian credible interval. Surgeons easily grasp this result and the visual display clearly illustrates how they perform compared to their peers on a quarterly basis. In addition, these risk-adjusted results allow surgeons to compare their patients´ outcomes with national benchmarks and to initiate quality improvement efforts as needed.

4a2.1.2. Describe the process(es) involved, including when/how often results were provided, what data were provided, what educational/explanatory efforts were made, etc.

Please see response under 4a2.1.1

4a2.2.1. Summarize the feedback on measure performance and implementation from the measured entities and others described in 4d.1.

Describe how feedback was obtained.

The adult cardiac surgeons from across the U.S. who comprise the STS Adult Cardiac Surgery Task Force meet periodically to discuss the participant reports and to consider potential enhancements to the ACSD. Additions/clarifications to the data collection form and to the content/format of the participant reports are discussed and implemented as appropriate.

Most recently, STS surgeon members have expressed interest in real-time, online data updates, which has led to the development of dashboard-type reporting on STS.org. Developed by IQVIA, the Society´s new data warehouse (https://www.sts.org/registries-research-center/sts-national-database/database-transition-resources), the new platform for the Adult Cardiac Surgery Database was released in early 2020. Surgeon members have access to near-real time data updates in the dashboard. Enhancements to dashboard functionality are ongoing.

Also, adult cardiac public reporting has been available since 2010 (http://publicreporting.sts.org/acsd), making star ratings for consenting participant groups available to participants as well as the public.

4a2.2.2. Summarize the feedback obtained from those being measured.

Please see response under 4a2.2.1

4a2.2.3. Summarize the feedback obtained from other users

Voluntary participation in ACSD public reporting has continually increased over the years that the initiative has been available, from 38% of ACSD participants in 2014, to 49% in 2016, to 67% in 2018, to approximately 79% in October 2020. This trend suggests that feedback from ACSD participants and others who access the performance data available on STS.org is sufficiently positive to promote ever-increasing participation in public reporting.

4a2.3. Describe how the feedback described in 4a2.2.1 has been considered when developing or revising the measure specifications or implementation, including whether the measure was modified and why or why not.

N/AN/A

Improvement

Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated. If not in use for performance improvement at the time of initial endorsement, then a credible rationale describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

4b1. Refer to data provided in 1b but do not repeat here. Discuss any progress on improvement (trends in performance results, number and percentage of people receiving high-quality healthcare; Geographic area and number and percentage of accountable entities and patients included.)

If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

The overall usage rates in the last three 12-month periods were 98.62%, 98.80% and 98.94% (January-December 2017, 2018 and 2019 respectively). This trend demonstrates the continuous progress on improvement that the STS expects to see in all our quality metrics.

Number of participants and operations by geographic regions, in January 2018 to December 2018 and in January 2019 to December 2019

Period January-December 2018

	Midwest	NE	Other*	South	West
# Part.	281	136	8	402	210
% Part.	27.1%	13.1%	0.8%	38.8%	20.3%
# Oper.	34325	24470	2864	66530	23616
% Oper.	22.6%	16.1%	1.9%	43.8%	15.6%

Period January-December 2019

	Midwest	NE	Other*	South	West
# Part.	263	134	1	392	209
% Part.	26.3%	13.4%	0.1%	39.2%	20.9%
# Oper.	33839	25467	5	67261	24201
% Oper.	22.444%	16.891%	0.003%	44.611%	16.051%

*Other: Ontario, Canada

If the above table is not clearly displayed, please refer to the version included in the appendix for this measure. The overall usage rates in the last three 12-month periods were 98.62%, 98.80% and 98.94% (January-December 2017, 2018 and 2019 respectively). This trend demonstrates the continuous progress on improvement that the STS expects to see in all our quality metrics.

Number of participants and operations by geographic regions, in January 2018 to December 2018 and in January 2019 to December 2019

Period January-December 2018

	Midwest	NE	Other*	South	West
# Part.	281	136	8	402	210
% Part.	27.1%	13.1%	0.8%	38.8%	20.3%
# Oper.	34325	24470	2864	66530	23616
% Oper.	22.6%	16.1%	1.9%	43.8%	15.6%

Period January-December 2019

	Midwest	NE	Other*	South	West
# Part.	263	134	1	392	209
% Part.	26.3%	13.4%	0.1%	39.2%	20.9%
# Oper.	33839	25467	5	67261	24201
% Oper.	22.444%	16.891%	0.003%	44.611%	16.051%

*Other: Ontario, Canada

If the above table is not clearly displayed, please refer to the version included in the appendix for this measure.

4b2. Unintended Consequences

The benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

4b2.1. Please explain any unexpected findings (positive or negative) during implementation of this measure including unintended impacts on patients.

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; 10% of STS Adult Cardiac Surgery Database participants were audited in each year from 2014 through 2019. (Our audit plans for 2020 were canceled due to the coronavirus pandemic; we expect to resume with 10% audits in 2021.) We control for risk aversion by having a robust methodology that appropriately adjusts the expected risk for providers who care for sicker patients.

4b2.2. Please explain any unexpected benefits from implementation of this measure.

N/AN/A

5. Comparison to Related or Competing Measures

If a measure meets the above criteria **and** there are endorsed or new related measures (either the same measure focus or the same target population) or competing measures (both the same measure focus and the same target population), the measures are compared to address harmonization and/or selection of the best measure.

5. Relation to Other NQF-endorsed Measures

Are there related measures (conceptually, either same measure focus or target population) or competing measures (conceptually both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.

Yes

5.1a. List of related or competing measures (selected from NQF-endorsed measures)

0114: Risk-Adjusted Postoperative Renal Failure

0115: Risk-Adjusted Surgical Re-exploration

0116: Anti-Platelet Medication at Discharge

0118: Anti-Lipid Treatment Discharge

0119: Risk-Adjusted Operative Mortality for CABG

0127: Preoperative Beta Blockade

0129: Risk-Adjusted Postoperative Prolonged Intubation (Ventilation)

0130: Risk-Adjusted Deep Sternal Wound Infection

0131: Risk-Adjusted Stroke/Cerebrovascular Accident

0134: Use of Internal Mammary Artery (IMA) in Coronary Artery Bypass Graft (CABG)

5.1b. If related or competing measures are not NQF endorsed, please indicate measure title and steward.

Additional related measure: 0696 - STS CABG Composite (not listed in drop-down menu for 5.1a)

5a. Harmonization of Related Measures

The measure specifications are harmonized with related measures.

OR

The differences in specifications are justified

5a.1. If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s):

Are the measure specifications harmonized to the extent possible?

Yes

5a.2. If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.

5b. Competing Measures

The measure is superior to competing measures (e.g., is a more valid or efficient way to measure).

OR

Multiple measures are justified.

5b.1. If this measure conceptually addresses both the same measure focus and the same target population as NQF-endorsed measure(s):

Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality); OR provide a rationale for the additive value of endorsing an additional measure. (Provide analyses when possible.)

N/A

Appendix

A.1 Supplemental materials may be provided in an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be organized in one file with a table of contents or bookmarks. If material pertains to a specific submission form number, that should be indicated. Requested information should be provided in the submission form and required attachments. There is no guarantee that supplemental materials will be reviewed.

Attachment Attachment: 0117_Beta_Blockade_at_Discharge_Appendix_-_S.9-_1b.2-_1b.4-_10212020-637407303665942835.pdf

Contact Information

Co.1 Measure Steward (Intellectual Property Owner): The Society of Thoracic Surgeons

Co.2 Point of Contact: Mark, Antman, mantman@sts.org, 312-202-5856-

Co.3 Measure Developer if different from Measure Steward: The Society of Thoracic Surgeons

Co.4 Point of Contact: Mark, Antman, mantman@sts.org, 312-202-5856-

Additional Information

Ad.1 Workgroup/Expert Panel involved in measure development

Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.

The STS Quality Measurement Task Force (chaired by David Shahian, MD) is responsible for measure development. Members of the STS Task Force on Quality Initiatives provide clinical expertise as needed. The STS Workforce on Quality meets at the STS Annual Meeting and reviews the measures on a yearly basis. Changes or updates to the measure will be at the recommendation of the Workforce.

Quality Measurement Task Force

David M. Shahian, MD, Chair; Massachusetts General Hospital & Harvard Medical School, Boston, MA

Diane Alejo; Johns Hopkins Univ., Baltimore, MD

Vinay Badhwar, MD; West Virginia University Hospitals, Morgantown, WV

Jordan Bloom, MD; Massachusetts General Hospital, Boston, MA

Michael Bowdish, MD; Torrance Memorial Medical Center, Los Angeles, CA

Joseph Cleveland, Jr., MD; University of Colorado Anschutz Medical Campus, Aurora, Co

Nimesh Desai, MD; Hospital of the University of Pennsylvania, Philadelphia, PA

James Edgerton, MD; Cardiac Surgery Specialists, Plano, TX

Fred Edwards, MD; University of Florida College of Medicine, Jacksonville, FL

Melanie Edwards, MD; Saint Joseph Mercy Health System, Ypsilanti, MI

Vic Ferraris, MD; University of Kentucky Medical Center, Lexington, KY

Anthony Furnary, MD; Providence Alaska Medical Center, Anchorage, AK

Joshua Goldberg, MD; Westchester Medical Center, Valhalla, NY

Jeffrey P. Jacobs, MD; University of Florida, Gainesville, FL

Marshall Jacobs, MD; Johns Hopkins Cardiac Surgery, Baltimore, MD

Karen Kim, MD; Univ. of Michigan Hospitals & Health Centers, Ann Arbor, MI

Benjamin Kozower, MD; Washington University School of Medicine, St. Louis, MO

Paul Kurlansky, MD; Columbia HeartSource/Columbia University Medical Center, New York, NY

Kevin Lobdell, MD; Atrium Health, Charlotte, NC

Mitchell Magee, MD; Southwest Cardiothoracic Surgeons, Dallas, TX

Gaetano Paone, MD; Henry Ford Hospital, Detroit, MI

J. Scott Rankin, MD; WVU Heart & Vascular Institute, West Virginia University, Morgantown, WV

Charles Schwartz, MD; St. Joseph Mercy Hospital, Pontiac, MI

Vinod Thourani, MD; MedStar Washington Hospital Center, Washington, DC

Christina Vassileva, MD; U Mass Memorial Medical Center, Worcester, MA

Moritz Wyler von Ballmoos, MD; Houston Methodist DeBakey Heart & Vascular Center, Houston, TX

Sean M. O'Brien, PhD; Duke Clinical Research Institute, Durham, NC The STS Quality Measurement Task Force (chaired by David Shahian, MD) is responsible for measure development. Members of the STS Task Force on Quality Initiatives provide clinical expertise as needed. The STS Workforce on Quality meets at the STS Annual Meeting and reviews the measures on a yearly basis. Changes or updates to the measure will be at the recommendation of the Workforce.

Quality Measurement Task Force

David M. Shahian, MD, Chair; Massachusetts General Hospital & Harvard Medical School, Boston, MA

Diane Alejo; Johns Hopkins Univ., Baltimore, MD

Vinay Badhwar, MD; West Virginia University Hospitals, Morgantown, WV

Jordan Bloom, MD; Massachusetts General Hospital, Boston, MA

Michael Bowdish, MD; Torrance Memorial Medical Center, Los Angeles, CA

Joseph Cleveland, Jr., MD; University of Colorado Anschutz Medical Campus, Aurora, Co

Nimesh Desai, MD; Hospital of the University of Pennsylvania, Philadelphia, PA

James Edgerton, MD; Cardiac Surgery Specialists, Plano, TX

Fred Edwards, MD; University of Florida College of Medicine, Jacksonville, FL

Melanie Edwards, MD; Saint Joseph Mercy Health System, Ypsilanti, MI

Vic Ferraris, MD; University of Kentucky Medical Center, Lexington, KY

Anthony Furnary, MD; Providence Alaska Medical Center, Anchorage, AK

Joshua Goldberg, MD; Westchester Medical Center, Valhalla, NY

Jeffrey P. Jacobs, MD; University of Florida, Gainesville, FL

Marshall Jacobs, MD; Johns Hopkins Cardiac Surgery, Baltimore, MD

Karen Kim, MD; Univ. of Michigan Hospitals & Health Centers, Ann Arbor, MI

Benjamin Kozower, MD; Washington University School of Medicine, St. Louis, MO

Paul Kurlansky, MD; Columbia HeartSource/Columbia University Medical Center, New York, NY

Kevin Lobdell, MD; Atrium Health, Charlotte, NC

Mitchell Magee, MD; Southwest Cardiothoracic Surgeons, Dallas, TX

Gaetano Paone, MD; Henry Ford Hospital, Detroit, MI

J. Scott Rankin, MD; WVU Heart & Vascular Institute, West Virginia University, Morgantown, WV

Charles Schwartz, MD; St. Joseph Mercy Hospital, Pontiac, MI

Vinod Thourani, MD; MedStar Washington Hospital Center, Washington, DC

Christina Vassileva, MD; U Mass Memorial Medical Center, Worcester, MA

Moritz Wyler von Ballmoos, MD; Houston Methodist DeBakey Heart & Vascular Center, Houston, TX

Sean M. O'Brien, PhD; Duke Clinical Research Institute, Durham, NC

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 Year the measure was first released: 2004

Ad.3 Month and Year of most recent revision: 06, 2016

Ad.4 What is your frequency for review/update of this measure? Annually

Ad.5 When is the next scheduled review/update for this measure? 01, 2021

Ad.6 Copyright statement: N/A

Ad.7 Disclaimers: N/A

Ad.8 Additional Information/Comments: N/A