



## Measure Information

This document contains the information submitted by measure developers/stewards, but is organized according to NQF's measure evaluation criteria and process. The item numbers refer to those in the submission form but may be in a slightly different order here. In general, the item numbers also reference the related criteria (e.g., item 1b.1 relates to sub criterion 1b).

## Brief Measure Information

**NQF #:** 0120

### Corresponding Measures:

**Measure Title:** Risk-Adjusted Operative Mortality for Aortic Valve Replacement (AVR)

**Measure Steward:** The Society of Thoracic Surgeons

**sp.02. Brief Description of Measure:** Percent of patients aged 18 years and older undergoing Aortic Valve Replacement (AVR) who die, including both 1) all deaths occurring during the hospitalization in which the procedure was performed, even if after 30 days, and 2) those deaths occurring after discharge from the hospital, but within 30 days of the procedure

**1b.01. Developer Rationale:** Mortality is likely the single most important negative outcome that can be associated with a surgical procedure. Operative mortality, defined as mortality within 30 days of surgery or on the same hospital admission, should include nearly all deaths that occur as a direct result of the surgery or an immediate postoperative complication. Critical evaluation of operative mortality allows one to evaluate the risk associated with a given procedure for various patient characteristics, and more importantly, aggressively search for ways to minimize that risk.

- Mihaljevic T, Nowicki ER, Rajeswaran J, et. al. Survival after valve replacement for aortic stenosis: implications for decision making. J Thorac Cardiovasc Surg. 2008 Jun;135(6):1270-8; discussion 1278-9.

- Tabata M, Umakanthan R, Cohn LH, et. al. Early and late outcomes of 1000 minimally invasive aortic valve operations. Eur J Cardiothorac Surg. 2008 Apr;33(4):537-41. Epub 2008 Feb 5.

- Chaliki HP, Mohty D, Avierinos JF, et. al. Outcomes after aortic valve replacement in patients with severe aortic regurgitation and markedly reduced left ventricular function. Circulation. 2002 Nov 19;106(21):2687-93

- Brennan JM, Holmes DR, Sherwood MW, Edwards FH, et al. The Association of Transcatheter Aortic Valve Replacement Availability and Hospital Aortic Valve Replacement Volume and Mortality in the United States. Ann Thorac Surg. 2014 Dec;98(6):2016-22.

- Thourani VH, Suri RM, et al. Contemporary real-world outcomes of surgical aortic valve replacement in 141,905 low-risk, intermediate-risk, and high-risk patients. Ann Thorac Surg. 2015 Jan;99(1):55-61.

- Patel HJ, Likosky DS, et al. Aortic Valve Replacement in the Moderately Elevated Risk Patient: A Population-Based Analysis of Outcomes. Ann Thorac Surg. 2016 Nov;102(5):1466-72.

- Johnston LE, Downs EA, et al. Outcomes for Low-Risk Surgical Aortic Valve Replacement: A Benchmark for Aortic Valve Technology. Ann Thorac Surg. 2017 Oct;104(4):1282-88.

- Chang C, Raza S, et al. Minimally Invasive Approaches to Surgical Aortic Valve Replacement: A Meta-Analysis. Ann Thorac Surg. 2018 Sep. [epub ahead of print]

---

**sp.12. Numerator Statement:** Number of patients aged 18 years and older undergoing AVR who die, including both 1) all deaths occurring during the hospitalization in which the operation was performed, even if after 30 days, and 2) those deaths occurring after discharge from the hospital, but within 30 days of the procedure

**sp.14. Denominator Statement:** All patients undergoing isolated AVR surgery

**sp.16. Denominator Exclusions:** N/A

---

**Measure Type:** Outcome

**sp.28. Data Source:**

Registry Data

**sp.07. Level of Analysis:**

Facility

Clinician: Group/Practice

---

**IF Endorsement Maintenance – Original Endorsement Date:** 2007-05-09 12:00 AM

**Most Recent Endorsement Date:** 6/10/2019 1:54:32 PM

---

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

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

**sp.03. IF PAIRED/GROUPED, what is the reason this measure must be reported with other measures to appropriately interpret results?:**

## 1. 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

**1ma.01. Indicate whether there is new evidence about the measure since the most recent maintenance evaluation. If yes, please briefly summarize the new evidence, and ensure you have updated entries in the Evidence section as needed.**

[Response Begins]

No

[Response Ends]

Please separate added or updated information from the most recent measure evaluation within each question response in the Importance to Measure and Report: Evidence section. For example:

**Current Submission:**

Updated evidence information here.

**Previous (Year) Submission:**

Evidence from the previous submission here.

**1a.01. Provide a logic model.**

*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.*

[Response Begins]

[Response Ends]

**1a.02. Provide evidence that the target population values the measured outcome, process, or structure and finds it meaningful.**

*Describe how and from whom input was obtained.*

[Response Begins]

[Response Ends]

**1a.03. Provide empirical data demonstrating the relationship between the outcome (or PRO) and at least one healthcare structure, process, intervention, or service.**

[Response Begins]

[Response Ends]

**1b.01. Briefly explain the rationale for this measure.**

*Explain how the measure will improve the quality of care, and list the benefits or improvements in quality envisioned by use of this measure.*

**[Response Begins]**

Mortality is likely the single most important negative outcome that can be associated with a surgical procedure. Operative mortality, defined as mortality within 30 days of surgery or on the same hospital admission, should include nearly all deaths that occur as a direct result of the surgery or an immediate postoperative complication. Critical evaluation of operative mortality allows one to evaluate the risk associated with a given procedure for various patient characteristics, and more importantly, aggressively search for ways to minimize that risk.

- Mihaljevic T, Nowicki ER, Rajeswaran J, et. al. Survival after valve replacement for aortic stenosis: implications for decision making. J Thorac Cardiovasc Surg. 2008 Jun;135(6):1270-8; discussion 1278-9.

- Tabata M, Umakanthan R, Cohn LH, et. al. Early and late outcomes of 1000 minimally invasive aortic valve operations. Eur J Cardiothorac Surg. 2008 Apr;33(4):537-41. Epub 2008 Feb 5.

- Chaliki HP, Mohty D, Avierinos JF, et. al. Outcomes after aortic valve replacement in patients with severe aortic regurgitation and markedly reduced left ventricular function. Circulation. 2002 Nov 19;106(21):2687-93

- Brennan JM, Holmes DR, Sherwood MW, Edwards FH, et al. The Association of Transcatheter Aortic Valve Replacement Availability and Hospital Aortic Valve Replacement Volume and Mortality in the United States. Ann Thorac Surg. 2014 Dec;98(6):2016-22.

- Thourani VH, Suri RM, et al. Contemporary real-world outcomes of surgical aortic valve replacement in 141,905 low-risk, intermediate-risk, and high-risk patients. Ann Thorac Surg. 2015 Jan;99(1):55-61.

- Patel HJ, Likosky DS, et al. Aortic Valve Replacement in the Moderately Elevated Risk Patient: A Population-Based Analysis of Outcomes. Ann Thorac Surg. 2016 Nov;102(5):1466-72.

- Johnston LE, Downs EA, et al. Outcomes for Low-Risk Surgical Aortic Valve Replacement: A Benchmark for Aortic Valve Technology. Ann Thorac Surg. 2017 Oct;104(4):1282-88.

- Chang C, Raza S, et al. Minimally Invasive Approaches to Surgical Aortic Valve Replacement: A Meta-Analysis. Ann Thorac Surg. 2018 Sep. [epub ahead of print]

**[Response Ends]**

**1b.02. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis.**

*Include mean, std dev, min, max, interquartile range, and 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 (4b) under Usability and Use.*

**[Response Begins]**

The measure was calculated using STS data for patients undergoing isolated AVR in two consecutive time periods, July 2011 – June 2014 and July 2014 – June 2017.

The summary statistic provided is the Participant's Estimated Odds Ratio (OR) based on a hierarchical logistic regression analysis. The OR measures the impact that a participant's performance level has on a patient's probability of experiencing an adverse outcome. An OR greater than 1.0 implies that the participant increases a

patient's risk of experiencing the outcome, relative to an average STS participant. An OR less than 1.0 implies that the participant decreases a patient's risk of experiencing the outcome, relative to an "average" STS participant. A high OR is undesirable and we define the percentiles with increasing OR. For example, 10% of STS participants have an OR greater than the value indicated by the "90th percentile" below.

Also provided is the distribution of the risk adjusted event rate. The risk adjusted rate is an estimate of the participant's event rate if, hypothetically, the case-mix of the patients treated by the participants is the same as the overall STS case-mix. It is calculated by the OR of the participant, other patient level parameter estimates from the hierarchical logistic model, and the overall STS event rate, by:

STS event rate \* (Participant's Expected Event Rate) / (Participant's Expected Event Rate Assuming Its Performance = STS Average Performance)

In the above equation, "Participant's Expected Event Rate" is calculated with the participant's actual OR, and "Participant's Expected Event Rate Assuming Its Performance = STS Average Performance" is calculated by assuming the participant's OR = 1 (i.e. no difference in performance from the STS average).

Distribution of participant-specific risk adjusted odds ratio and event rates in July 2011 - June 2014 and July 2014 - June 2017

Distribution July 2011 - June 2014 Odds ratio July 2011 - June 2014 Risk adjusted Rate, % July 2014 - June 2017

Odds ratio July 2014 - June 2017 Risk adjusted Rate, %

# Participant 1085 1085 1093 1093

# Operations 87333 87333 86163 86163

Mean 1.08 2.62 1.04 2.27

STD 0.39 0.86 0.24 0.48

IQR 0.35 0.83 0.24 0.50

0% 0.24 0.62 0.49 1.09

10% 0.73 1.83 0.80 1.78

20% 0.83 2.05 0.87 1.92

30% 0.88 2.19 0.92 2.03

40% 0.94 2.31 0.96 2.11

50% 0.98 2.41 0.99 2.17

60% 1.03 2.53 1.03 2.26

70% 1.15 2.81 1.10 2.41

80% 1.29 3.12 1.18 2.58

90% 1.51 3.62 1.33 2.89

100% 4.79 9.98 3.11 5.81

Midwest 314 314 305 305

Northeast 138 138 142 142

Other\* 0 0 9 9

South 408 408 415 415

West 225 225 222 222

\*Other Region: Ontario, Canada

(Please see Appendix if table of performance values does not display clearly above.)

[Response Ends]

**1b.03. If no or limited performance data on the measure as specified is reported above, 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. Include citations.**

[Response Begins]

N/A

**[Response Ends]**

**1b.04. 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.**

*Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included. Include mean, std dev, min, max, interquartile range, and scores by decile. 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 (4b) under Usability and Use.*

**[Response Begins]**

Even though the measure is used to measure participant-level results, we understand it is of interest to see whether disparity exists between race and sex groups. We provide below the participant level distribution of the measure by race, ethnicity and sex.

Distribution of participant-specific risk adjusted odds ratio in July 2011 - June 2014 and July 2014 - June 2017, by sex

Distribution Male July 11 - June 14 Male July 14 - June 17 Female July 11 - June 14 Female July 14 - June 17

# Participant 1075 1086 1065 1076

# Operations 52274 53179 35059 32984

Mean 1.06 1.04 1.07 1.03

STD 0.32 0.24 0.35 0.22

IQR 0.30 0.25 0.32 0.21

0% 0.28 0.45 0.38 0.56

10% 0.78 0.82 0.77 0.84

20% 0.85 0.88 0.85 0.89

30% 0.90 0.92 0.89 0.92

40% 0.94 0.95 0.93 0.95

50% 0.97 0.97 0.97 0.97

60% 0.99 0.99 0.99 0.99

70% 1.12 1.08 1.12 1.08

80% 1.26 1.20 1.26 1.16

90% 1.50 1.34 1.47 1.30

100% 3.11 2.72 4.43 2.64

Distribution of participant-specific risk adjusted event rates (%) in July 2011 - June 2014 and July 2014 - June 2017, by sex

Distribution Male July 11 - June 14 Male July 14 - June 17 Female July 11 - June 14 Female July 14 - June 17

# Participant 1075 1086 1065 1076

# Operations 52274 53179 35059 32984

Mean 2.22 1.86 3.15 2.93

STD 0.62 0.41 0.96 0.57

IQR 0.61 0.43 0.91 0.57

0% 0.60 0.83 1.17 1.63

10% 1.67 1.48 2.33 2.42

20% 1.81 1.60 2.54 2.55

30% 1.91 1.65 2.66 2.65

40% 1.97 1.71 2.78 2.72

50% 2.04 1.75 2.88 2.78

#0120 Risk-Adjusted Operative Mortality for Aortic Valve Replacement (AVR), Submission Last Updated:  
Sep 09, 2022

60% 2.10 1.79 2.96 2.84  
70% 2.35 1.94 3.32 3.05  
80% 2.61 2.13 3.70 3.28  
90% 3.06 2.37 4.27 3.63  
100% 6.11 4.70 11.97 7.12

Distribution of participant-specific risk adjusted odds ratio in July 2011 - June 2014 and July 2014 - June 2017, by age

Distribution Age < 75, July 11 - June 14 Age < 75, July 14- June 17 Age = 75, July 11 - June 14 Age = 75, July 14 - June 17

# Participant 1081 1090 1041 1057

# Operations 55200 61665 32133 24498

Mean 1.06 1.04 1.05 1.02

STD 0.33 0.24 0.29 0.14

IQR 0.28 0.25 0.29 0.16

0% 0.30 0.48 0.38 0.65

10% 0.79 0.82 0.79 0.88

20% 0.86 0.87 0.86 0.92

30% 0.90 0.92 0.91 0.94

40% 0.94 0.95 0.94 0.96

50% 0.97 0.98 0.97 0.98

60% 1.00 1.01 1.00 0.99

70% 1.11 1.10 1.11 1.05

80% 1.24 1.18 1.24 1.12

90% 1.39 1.32 1.43 1.19

100% 4.62 3.96 2.83 1.95

Distribution of participant-specific risk adjusted event rates (%) in July 2011 - June 2014 and July 2014 - June 2017, by age

Distribution Age < 75, July 11 - June 14 Age < 75, July 14- June 17 Age = 75, July 11 - June 14 Age = 75, July 14 - June 17

# Participant 1081 1090 1041 1057

# Operations 55200 61665 32133 24498

Mean 2.02 1.97 3.52 2.98

STD 0.58 0.42 0.90 0.40

IQR 0.51 0.45 0.94 0.45

0% 0.58 0.93 1.32 1.95

10% 1.53 1.57 2.69 2.61

20% 1.67 1.67 2.93 2.72

30% 1.74 1.75 3.07 2.78

40% 1.80 1.81 3.18 2.84

50% 1.87 1.86 3.27 2.89

60% 1.92 1.92 3.35 2.92

70% 2.12 2.08 3.72 3.09

80% 2.34 2.24 4.12 3.26

90% 2.64 2.47 4.66 3.47

100% 7.58 5.86 9.06 5.55

At the operation level, we were able to estimate the risk adjusted odds ratios between race groups. The odds ratios were estimated from a model with race and other covariates from the 2008 validated Valve risk models.

Risk Adjusted OR:

- Black vs. White (including patients with race other than Black, White, Asian):

1.04 (0.85-1.26)

- Asian vs. White (including patients with race other than Black, White, Asian):

1.05 (0.73-1.50)

(Please see Appendix if tables do not display clearly above.)

**[Response Ends]**

**1b.05. If no or limited data on disparities from the measure as specified is reported above, 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 above.**

**[Response Begins]**

N/A

**[Response Ends]**



## 2. Scientific Acceptability of Measure Properties

Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. 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.

**spma.01. Indicate whether there are changes to the specifications since the last updates/submission. If yes, update the specifications in the Measure Specifications section of the Measure Submission Form, and explain your reasoning for the changes below.**

[Response Begins]

No

[Response Ends]

**spma.02. Briefly describe any important changes to the measure specifications since the last measure update and provide a rationale.**

For annual updates, please explain how the change in specifications affects the measure results. If a material change in specification is identified, data from re-testing of the measure with the new specifications is required for early maintenance review.

*For example, specifications may have been updated based on suggestions from a previous NQF CDP review.*

[Response Begins]

No changes since last measure update.

[Response Ends]

**sp.01. Provide the measure title.**

*Measure titles should be concise yet convey who and what is being measured (see [What Good Looks Like](#)).*

[Response Begins]

Risk-Adjusted Operative Mortality for Aortic Valve Replacement (AVR)

[Response Ends]

**sp.02. Provide a brief description of the measure.**

*Including type of score, measure focus, target population, timeframe, (e.g., Percentage of adult patients aged 18-75 years receiving one or more HbA1c tests per year).*

[Response Begins]

Percent of patients aged 18 years and older undergoing Aortic Valve Replacement (AVR) who die, including both 1) all deaths occurring during the hospitalization in which the procedure was performed, even if after 30 days, and 2) those deaths occurring after discharge from the hospital, but within 30 days of the procedure

[Response Ends]

**sp.04. Check all the clinical condition/topic areas that apply to your measure, below.**

*Please refrain from selecting the following answer option(s). We are in the process of phasing out these answer options and request that you instead select one of the other answer options as they apply to your measure.*

*Please do not select:*

- *Surgery: General*

**[Response Begins]**

Cardiovascular

Surgery

Surgery: Cardiac Surgery

**[Response Ends]**

**sp.05. Check all the non-condition specific measure domain areas that apply to your measure, below.**

**[Response Begins]**

Safety

Safety: Complications

**[Response Ends]**

**sp.06. Select one or more target population categories.**

*Select only those target populations which can be stratified in the reporting of the measure's result.*

*Please refrain from selecting the following answer option(s). We are in the process of phasing out these answer options and request that you instead select one of the other answer options as they apply to your measure.*

*Please do not select:*

- *Populations at Risk: Populations at Risk*

**[Response Begins]**

Adults (Age >= 18)

**[Response Ends]**

**sp.07. Select the levels of analysis that apply to your measure.**

*Check ONLY the levels of analysis for which the measure is SPECIFIED and TESTED.*

*Please refrain from selecting the following answer option(s). We are in the process of phasing out these answer options and request that you instead select one of the other answer options as they apply to your measure.*

*Please do not select:*

- *Clinician: Clinician*
- *Population: Population*

**[Response Begins]**

Clinician: Group/Practice

Facility

**[Response Ends]**

**sp.08. Indicate the care settings that apply to your measure.**

*Check ONLY the settings for which the measure is SPECIFIED and TESTED.*

**[Response Begins]**

Inpatient/Hospital

**[Response Ends]**

**sp.09. 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. If no URL is available, indicate "none available".*

**[Response Begins]**

[https://www.sts.org/sites/default/files/documents/STSAAdultCVDDataCollectionFormV4\\_20\\_2\\_ChangesHighlightedGOLDEN12212020Annotated.pdf](https://www.sts.org/sites/default/files/documents/STSAAdultCVDDataCollectionFormV4_20_2_ChangesHighlightedGOLDEN12212020Annotated.pdf)

[https://www.sts.org/sites/default/files/ACSD\\_DataSpecifications\\_V4\\_20\\_2.pdf](https://www.sts.org/sites/default/files/ACSD_DataSpecifications_V4_20_2.pdf)

**[Response Ends]**

**sp.12. Attach the data dictionary, code table, or value sets (and risk model codes and coefficients when applicable). Excel formats (.xlsx or .csv) are preferred.**

*Attach an excel or csv file; if this poses an issue, [contact staff](#). Provide descriptors for any codes. Use one file with multiple worksheets, if needed.*

**[Response Begins]**

No data dictionary/code table – all information provided in the submission form

**[Response Ends]**

For the question below: state the outcome being measured. Calculation of the risk-adjusted outcome should be described in sp.22.

**sp.13. State the numerator.**

*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.*

**[Response Begins]**

Number of patients aged 18 years and older undergoing AVR who die, including both 1) all deaths occurring during the hospitalization in which the operation was performed, even if after 30 days, and 2) those deaths occurring after discharge from the hospital, but within 30 days of the procedure

**[Response Ends]**

For the question below: describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in sp.22.

**sp.14. Provide details needed to calculate the numerator.**

*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 sp.11.*

**[Response Begins]**

Number of isolated AVR procedures in which Mortality Operative Death (MtOpD) is marked “yes.” Operative mortality is further verified by the following variables: Mortality Status at 30 days (Mt30Stat), Mortality Date (MtDate), (MtDCStat is Dead in version 2,81 or DischMortStat is Died in Hospital in version 2.9)

Version 4.20.2

Number of isolated AVR procedures in which Mortality Operative Death (MtOpD) is marked “yes.” Operative mortality is further verified by the following variables: Mortality Status at 30 days (Mt30Stat), Mortality Date (MtDate), Patient Expired in the OR (ExpiredInOR), Discharge status (DischMortStat) is Discharged to Hospice OR Died in Hospital

**[Response Ends]**

For the question below: state the target population for the outcome. Calculation of the risk-adjusted outcome should be described in sp.22.

**sp.15. State the denominator.**

*Brief, narrative description of the target population being measured.*

**[Response Begins]**

All patients undergoing isolated AVR surgery

**[Response Ends]**

For the question below: describe how the target population is identified. Calculation of the risk-adjusted outcome should be described in sp.22.

**sp.16. Provide details needed to calculate the denominator.**

*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 sp.11.*

**[Response Begins]**

Number of isolated AVR procedures.

**[Response Ends]**

**sp.17. Describe the denominator exclusions.**

*Brief narrative description of exclusions from the target population.*

**[Response Begins]**

N/A

**[Response Ends]**

**sp.18. Provide details needed to calculate the denominator exclusions.**

*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 sp.11.*

**[Response Begins]**

N/A

**[Response Ends]**

**sp.19. Provide all information required to stratify the measure results, if necessary.**

*Include 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 in the Data Dictionary field.*

**[Response Begins]**

N/A

**[Response Ends]**

**sp.20. Is this measure adjusted for socioeconomic status (SES)?**

**[Response Begins]**

Yes

**[Response Ends]**

**sp.21. Select the risk adjustment type.**

*Select type. Provide specifications for risk stratification and/or risk models in the Scientific Acceptability section.*

**[Response Begins]**

Statistical risk model

**[Response Ends]**

**sp.22. Select the most relevant type of score.**

*Attachment: If available, please provide a sample report.*

**[Response Begins]**

Rate/proportion

**[Response Ends]**

**sp.23. Select the appropriate interpretation of the measure score.**

*Classifies interpretation of score according to whether better quality or resource use is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score*

**[Response Begins]**

Better quality = Lower score

**[Response Ends]**

**sp.24. Diagram or describe the calculation of the measure score as an ordered sequence of steps.**

*Identify the target population; exclusions; cases meeting the target process, condition, event, or outcome; time period of data, aggregating data; risk adjustment; etc.*

**[Response Begins]**

Please refer to numerator and denominator sections for detailed information.

**[Response Ends]**

**sp.27. If measure testing is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.**

*Examples of samples used for testing:*

- Testing may be conducted on a sample of the accountable entities (e.g., hospital, physician). The analytic unit specified for the particular measure (e.g., physician, hospital, home health agency) determines the sampling strategy for scientific acceptability testing.
- The sample should represent the variety of entities whose performance will be measured. The [2010 Measure Testing Task Force](#) recognized that the samples used for reliability and validity testing often have limited generalizability because measured entities volunteer to participate. Ideally, however, all types of entities whose performance will be measured should be included in reliability and validity testing.
- The sample should include adequate numbers of units of measurement and adequate numbers of patients to answer the specific reliability or validity question with the chosen statistical method.
- When possible, units of measurement and patients within units should be randomly selected.

**[Response Begins]**

N/A

**[Response Ends]**

**sp.30. Select only the data sources for which the measure is specified.**

**[Response Begins]**

Registry Data

**[Response Ends]**

**sp.31. Identify the specific data source or data collection instrument.**

*For example, provide the name of the database, clinical registry, collection instrument, etc., and describe how data are collected.*

**[Response Begins]**

STS Adult Cardiac Surgery Database Version 2.9 (effective July 1, 2017); and Version 4.20 (effective July 1, 2020).

**[Response Ends]**

**sp.32. Provide the data collection instrument.**

**[Response Begins]**

Available at measure-specific web page URL identified in sp.09

**[Response Ends]**

**2ma.01. Indicate whether additional empirical reliability testing at the accountable entity level has been conducted. If yes, please provide results in the following section, Scientific Acceptability: Reliability - Testing. Include information on all testing conducted (prior testing as well as any new testing).**

*Please separate added or updated information from the most recent measure evaluation within each question response in the Scientific Acceptability sections. For example:*

**Current Submission:**

*Updated testing information here.*

**Previous Submission:**

*Testing from the previous submission here.*

**[Response Begins]**

Yes

**[Response Ends]**

**2ma.02. Indicate whether additional empirical validity testing at the accountable entity level has been conducted. If yes, please provide results in the following section, Scientific Acceptability: Validity - Testing. Include information on all testing conducted (prior testing as well as any new testing).**

*Please separate added or updated information from the most recent measure evaluation within each question response in the Scientific Acceptability sections. For example:*

**Current Submission:**

*Updated testing information here.*

**Previous Submission:**

*Testing from the previous submission here.*

**[Response Begins]**

Yes

**[Response Ends]**

**2ma.03. For outcome, patient-reported outcome, resource use, cost, and some process measures, risk adjustment/stratification may be conducted. Did you perform a risk adjustment or stratification analysis?**

**[Response Begins]**

Yes

**[Response Ends]**

**2ma.04. For maintenance measures in which risk adjustment/stratification has been performed, indicate whether additional risk adjustment testing has been conducted since the most recent maintenance evaluation. This may include updates to the risk adjustment analysis with additional clinical, demographic, and social risk factors.**

**Please update the Scientific Acceptability: Validity - Other Threats to Validity section.**

**Note: This section must be updated even if social risk factors are not included in the risk adjustment strategy.**

**[Response Begins]**

**[Response Ends]**

Measure testing must demonstrate adequate reliability and validity in order to be recommended for endorsement. Testing may be conducted for data elements and/or the computed measure score. Testing information and results should be entered in the appropriate fields in the Scientific Acceptability sections of the Measure Submission Form.

- Measures must be tested for all the data sources and levels of analyses that are specified. If there is more than one set of data specifications or more than one level of analysis, contact NQF staff about how to present all the testing information in one form.
- All required sections must be completed.
- For composites with outcome and resource use measures, Questions 2b.23-2b.37 (Risk Adjustment) also must be completed.
- If specified for multiple data sources/sets of specifications (e.g., claims and EHRs), Questions 2b.11-2b.13 also must be completed.
- An appendix for supplemental materials may be submitted (see Question 1 in the Additional section), but there is no guarantee it will be reviewed.
- Contact NQF staff with any questions. Check for resources at the [Submitting Standards webpage](#).



- For information on the most updated guidance on how to address social risk factors variables and testing in this form refer to the release notes for the [2021 Measure Evaluation Criteria and Guidance](#).

Note: The information provided in this form is intended to aid the Standing Committee and other stakeholders in understanding to what degree the testing results for this measure meet NQF's evaluation criteria for testing.

2a. Reliability testing demonstrates 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. For instrument-based measures (including PRO-PMs) and composite performance measures, reliability should be demonstrated for the computed performance score.

2b1. Validity testing demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For instrument based measures (including PRO-PMs) and composite performance measures, validity should be demonstrated for the computed performance score.

2b2. Exclusions are supported by the clinical evidence and are of sufficient frequency to warrant inclusion in the specifications of the measure;

AND

If patient preference (e.g., informed decision-making) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately).

2b3. For outcome measures and other measures when indicated (e.g., resource use):

- an evidence-based risk-adjustment strategy (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and social risk factors) that influence the measured outcome and are present at start of care; 14,15 and has demonstrated adequate discrimination and calibration

OR

- rationale/data support no risk adjustment/ stratification.

2b4. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for identification of statistically significant and practically/clinically meaningful 16 differences in performance;

OR

there is evidence of overall less-than-optimal performance.

2b5. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b6. Analyses 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.

2c. For composite performance measures, empirical analyses support the composite construction approach and demonstrate that:

2c1. the component measures fit the quality construct and add value to the overall composite while achieving the related objective of parsimony to the extent possible; and

2c2. the aggregation and weighting rules are consistent with the quality construct and rationale while achieving the related objective of simplicity to the extent possible.

(if not conducted or results not adequate, justification must be submitted and accepted)

## Definitions

Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).

Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measure scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality. The degree of consensus and any areas of disagreement must be provided/discussed.

Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.

Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

Risk factors that influence outcomes should not be specified as exclusions.

With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of \$25 in cost for an episode of care (e.g., \$5,000 v. \$5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

Please separate added or updated information from the most recent measure evaluation within each question response in the Scientific Acceptability sections. For example:

**Current Submission:**

Updated testing information here.

**Previous (Year) Submission:**

Testing from the previous submission here.

**2a.01. Select only the data sources for which the measure is tested.**

**[Response Begins]**

Registry Data

**[Response Ends]**

**2a.02. 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).*

**[Response Begins]**

STS Adult Cardiac Surgery Database Version 2.9 (effective July 1, 2017); and Version 4.20 (effective July 1, 2020).

[Response Ends]

**2a.03. Provide the dates of the data used in testing.**

*Use the following format: "MM-DD-YYYY - MM-DD-YYYY"*

[Response Begins]

[Response Ends]

**2a.04. Select the levels of analysis for which the measure is tested.**

*Testing must be provided for all the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan.*

*Please refrain from selecting the following answer option(s). We are in the process of phasing out these answer options and request that you instead select one of the other answer options as they apply to your measure.*

*Please do not select:*

- Clinician: Clinician
- Population: Population

[Response Begins]

Clinician: Group/Practice

Facility

[Response Ends]

**2a.05. List the measured entities 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.*

[Response Begins]

[Response Ends]

**2a.06. Identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis), separated by level of analysis and data source; if a sample was used, describe how patients were selected for inclusion in the sample.**

*If there is a minimum case count used for testing, that minimum must be reflected in the specifications.*

[Response Begins]

[Response Ends]

**2a.07. 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.**

[Response Begins]

[Response Ends]

**2a.08. List 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.*

**[Response Begins]**

**[Response Ends]**

Note: If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – in 2a.09 check patient or encounter-level data; in 2a.010 enter “see validity testing section of data elements”; and enter “N/A” for 2a.11 and 2a.12.

**2a.09. Select the level of reliability testing conducted.**

*Choose one or both levels.*

**[Response Begins]**

**[Response Ends]**

**2a.10. For each level of reliability testing 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.*

**[Response Begins]**

**[Response Ends]**

**2a.11. For each level of reliability testing checked above, what were the statistical results from reliability testing?**

*For example, provide the percent agreement and kappa for the critical data elements, or distribution of reliability statistics from a signal-to-noise analysis. For score-level reliability testing, when using a signal-to-noise analysis, more than just one overall statistic should be reported (i.e., to demonstrate variation in reliability across providers). If a particular method yields only one statistic, this should be explained. In addition, reporting of results stratified by sample size is preferred (pg. 18, [NQF Measure Evaluation Criteria](#)).*

**[Response Begins]**

**[Response Ends]**

**2a.12. Interpret the results, in terms of how they demonstrate reliability.**

*(In other words, what do the results mean and what are the norms for the test conducted?)*

**[Response Begins]**

**[Response Ends]**

**2b.01. Select the level of validity testing that was conducted.**

[Response Begins]

[Response Ends]

**2b.02. 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.*

[Response Begins]

[Response Ends]

**2b.03. Provide the statistical results from validity testing.**

*Examples may include correlations or t-test results.*

[Response Begins]

[Response Ends]

**2b.04. Provide 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?)**

[Response Begins]

[Response Ends]

**2b.05. 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 in Importance to Measure and Report: Gap in Care/Disparities.*

[Response Begins]

[Response Ends]

**2b.06. Describe the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities.**

*Examples may include 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.*

[Response Begins]

[Response Ends]

**2b.07. Provide 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.**

*In other words, what do the results mean in terms of statistical and meaningful differences?*

[Response Begins]

[Response Ends]

**2b.08. Describe the method of testing conducted to identify the extent and distribution of missing data (or non-response) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and non-responders). Include how the specified handling of missing data minimizes bias.**

*Describe the steps—do not just name a method; what statistical analysis was used.*

[Response Begins]

[Response Ends]

**2b.09. Provide the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data.**

*For example, provide results of sensitivity analysis of the effect of various rules for missing data/non-response. If no empirical sensitivity analysis was conducted, identify the approaches for handling missing data that were considered and benefits and drawbacks of each).*

[Response Begins]

[Response Ends]

**2b.10. Provide 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.**

*In other words, 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 was conducted, justify the selected approach for missing data.*

[Response Begins]

[Response Ends]

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 eQMs). 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.

**2b.11. Indicate whether there is more than one set of specifications for this measure.**

[Response Begins]

[Response Ends]

**2b.12. 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. Indicate what statistical analysis was used.*

[Response Begins]

[Response Ends]

**2b.13. Provide the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications.**

*Examples may include correlation, and/or rank order.*

[Response Begins]

[Response Ends]

**2b.14. Provide your interpretation of the results in terms of the differences in performance measure scores for the same entities across the different data sources/specifications.**

*In other words, what do the results mean and what are the norms for the test conducted.*

[Response Begins]

[Response Ends]

**2b.15. Indicate whether the measure uses exclusions.**

[Response Begins]

[Response Ends]

**2b.16. Describe the method of testing exclusions and what was tested.**

*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?*

[Response Begins]

[Response Ends]

**2b.17. Provide 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.*

[Response Begins]

[Response Ends]

**2b.18. Provide your interpretation of the results, in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results.**

*In other words, 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.*

[Response Begins]

[Response Ends]

**2b.19. Check all methods used to address risk factors.**

[Response Begins]

[Response Ends]

**2b.20. If using statistical risk models, provide detailed risk model specifications, including the risk model method, risk factors, risk factor data sources, coefficients, equations, codes with descriptors, and definitions.**

[Response Begins]

[Response Ends]

**2b.21. If an outcome or resource use measure is not risk-adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (i.e., case mix) is not needed to achieve fair comparisons across measured entities.**

[Response Begins]

[Response Ends]

**2b.22. Select all applicable resources and methods used to develop the conceptual model of how social risk impacts this outcome.**

[Response Begins]

[Response Ends]

**2b.23. Describe the conceptual and statistical methods and criteria used to test and select patient-level risk factors (e.g., clinical factors, social risk factors) used in the statistical risk model or for stratification by risk.**

*Please be sure to address the following: potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of  $p < 0.10$  or other statistical tests; correlation of  $x$  or higher. Patient factors should be present at the start of care, if applicable. Also discuss any "ordering" of risk factor inclusion; note whether social risk factors are added after all clinical factors. Discuss any considerations regarding data sources (e.g., availability, specificity).*

[Response Begins]

[Response Ends]

**2b.24. Detail the statistical results of the analyses used to test and select risk factors for inclusion in or exclusion from the risk model/stratification.**



[Response Begins]

[Response Ends]

**2b.25. Describe the analyses and interpretation resulting in the decision to select or not select social risk factors.**

*Examples may include prevalence of the factor across measured entities, availability of the data source, empirical association with the outcome, contribution of unique variation in the outcome, or assessment of between-unit effects and within-unit effects. Also describe the impact of adjusting for risk (or making no adjustment) on providers at high or low extremes of risk.*

[Response Begins]

[Response Ends]

**2b.26. 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 control for differences in patient characteristics (i.e., case mix) below. If stratified ONLY, enter “N/A” for questions about the statistical risk model discrimination and calibration statistics.**

*Validation testing should be conducted in a data set that is separate from the one used to develop the model.*

[Response Begins]

[Response Ends]

**2b.27. Provide risk model discrimination statistics.**

*For example, provide c-statistics or R-squared values.*

[Response Begins]

[Response Ends]

**2b.28. Provide the statistical risk model calibration statistics (e.g., Hosmer-Lemeshow statistic).**

[Response Begins]

[Response Ends]

**2b.29. Provide the risk decile plots or calibration curves used in calibrating the statistical risk model.**

*The preferred file format is .png, but most image formats are acceptable.*

[Response Begins]

[Response Ends]

**2b.30. Provide the results of the risk stratification analysis.**

[Response Begins]

[Response Ends]

**2b.31. Provide your interpretation of the results, in terms of demonstrating adequacy of controlling for differences in patient characteristics (i.e., case mix).**

*In other words, what do the results mean and what are the norms for the test conducted?*

**[Response Begins]**

**[Response Ends]**

**2b.32. Describe any additional testing conducted to justify the risk adjustment approach used in specifying the measure.**

*Not required but would provide additional support of adequacy of the risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed.*

**[Response Begins]**

**[Response Ends]**

### 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.

---

**3.01. Check all methods below that are used to generate the data elements needed to compute the measure score.**

**[Response Begins]**

Abstracted from a record by someone other than person obtaining original information (e.g., chart abstraction for quality measure or registry)

**[Response Ends]**

**3.02. Detail to what extent the specified data elements are available electronically in defined fields.**

*In other words, indicate whether data elements that are needed to compute the performance measure score are in defined, computer-readable fields.*

**[Response Begins]**

Some data elements are in defined fields in electronic sources

**[Response Ends]**

**3.03. 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 data elements not from electronic sources.**

**[Response Begins]**

As of November 2018, the STS Adult Cardiac Surgery Database has 1,091 participants in the U.S. and Canada, 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 Adult Cardiac Surgery Database 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 Adult Cardiac Surgery Database data elements.

**[Response Ends]**

**3.04. Describe any efforts to develop an eCQM.**

**[Response Begins]**

**[Response Ends]**

**3.06. 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.**

**[Response Begins]**

**[Response Ends]**

Consider implications for both individuals providing data (patients, service recipients, respondents) and those whose performance is being measured.

**3.07. Detail 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),**

**Attach the fee schedule here, if applicable.**

**[Response Begins]**

Data Collection:

There are no direct costs to collect the data for this measure. Costs to develop 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 (single cardiothoracic surgeons or 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.

**[Response Ends]**

## 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.

---

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

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 demonstrating performance improvement.

### 4a.01. Check all current uses. For each current use checked, please provide:

Name of program and sponsor

URL

Purpose

Geographic area and number and percentage of accountable entities and patients included

Level of measurement and setting

[Response Begins]

Public Reporting

[Response Ends]

### 4a.02. Check all planned uses.

[Response Begins]

[Response Ends]

### 4a.03. If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing), explain why the measure is not in use.

*For example, do policies or actions of the developer/steward or accountable entities restrict access to performance results or block implementation?*

[Response Begins]

N/A

[Response Ends]

### 4a.04. If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes: used in any accountability application within 3 years, and publicly reported within 6 years of initial endorsement.

*A 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.*

[Response Begins]

N/A

**[Response Ends]**

**4a.05. Describe how performance results, data, and assistance with interpretation have been provided to those being measured or other users during development or implementation.**

*Detail 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.*

**[Response Begins]**

As of November 2018, there are 1,091 active U.S. and Canadian participants in the STS Adult Cardiac Surgery Database (ACSD). A "participant" is a cardiothoracic surgeon or 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 surgeon or 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.

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.

**[Response Ends]**

**4a.06. Describe the process for providing measure results, including when/how often results were provided, what data were provided, what educational/explanatory efforts were made, etc.**

**[Response Begins]**

Please see response under 4a2.1.1

**[Response Ends]**

**4a.07. Summarize the feedback on measure performance and implementation from the measured entities and others. Describe how feedback was obtained.**

**[Response Begins]**

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. Roll-out of the adult cardiac dashboard is underway in 2018.

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.

**[Response Ends]**

**4a.08. Summarize the feedback obtained from those being measured.**

**[Response Begins]**

Please see response under 4a2.2.1

**[Response Ends]**

**4a.09. Summarize the feedback obtained from other users.**

**[Response Begins]**

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 approximately 67% as of November 2018. 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.

**[Response Ends]**

**4a.10. Describe how the feedback described has been considered when developing or revising the measure specifications or implementation, including whether the measure was modified and why or why not.**

**[Response Begins]**

N/A

**[Response Ends]**

**4b.01. You may refer to data provided in Importance to Measure and Report: Gap in Care/Disparities, 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, provide an explanation. 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.**

**[Response Begins]**

Looking at the overall temporal trend, the operative mortality rate has been steadily declining. The overall event rates in the last three 12-month periods were 2.34%, 2.17%, 2.09% (July 2014-June 2015, July 2015-June 2016, July 2016-June 2017, respectively).

Number of participants and operations by geographic regions, in July 2011 to June 2014 and in July 2014 to June 2017

July 2011 to June 2014 July 2014 to June 2017

Midwest Northeast Other region\* South West Midwest Northeast Other region\* South West

# Participant 314 138 0 408 225 # Participant 305 142 9 415 222

% Participant 28.9% 12.7% 0.0% 37.6% 20.7% % Participant 27.9% 13.0% 0.8% 38.0% 20.3%

# Operation 21479 19239 NA 28312 18303 # Operation 20831 18767 578 28755 17232

% Operation 24.6% 22.0% NA% 32.4% 21.0% % Operation 24.2% 21.8% 0.7% 33.4% 20.0%

\*Other Region: Ontario, Canada

(Please see Appendix if table does not display clearly above.)

**[Response Ends]**

**4b.02. Explain any unexpected findings (positive or negative) during implementation of this measure, including unintended impacts on patients.**

**[Response Begins]**

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 2018, as in each year since 2014. We control for risk aversion by having a robust methodology that appropriately adjusts the expected risk for providers who care for sicker patients.

**[Response Ends]**

**4b.03. Explain any unexpected benefits realized from implementation of this measure.**

**[Response Begins]**

**[Response Ends]**



## 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.

---

If you are updating a maintenance measure submission for the first time in MIMS, please note that the previous related and competing data appearing in question 5.03 may need to be entered in to 5.01 and 5.02, if the measures are NQF endorsed. Please review and update questions 5.01, 5.02, and 5.03 accordingly.

### 5.01. Search and select all NQF-endorsed related measures (conceptually, either same measure focus or target population).

*(Can search and select measures.)*

[Response Begins]

[Response Ends]

### 5.02. Search and select all NQF-endorsed competing measures (conceptually, the measures have both the same measure focus or target population).

*(Can search and select measures.)*

[Response Begins]

[Response Ends]

### 5.03. If there are related or competing measures to this measure, but they are not NQF-endorsed, please indicate the measure title and steward.

[Response Begins]

Also related and NQF-endorsed (not available with search function in 5.1a):

2561 STS Aortic Valve Replacement (AVR) Composite Score

[Response Ends]

### 5.04. If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s), indicate whether the measure specifications are harmonized to the extent possible.

[Response Begins]

Yes

[Response Ends]

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

[Response Begins]

N/A

[Response Ends]

### 5.06. Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality). Alternatively, justify endorsing an additional measure.

*Provide analyses when possible.*

**[Response Begins]**

N/A

**[Response Ends]**

## Appendix

Supplemental materials may be provided in an appendix.:

## Contact Information

**Measure Steward (Intellectual Property Owner):** The Society of Thoracic Surgeons

**Measure Steward Point of Contact:** Yagci, Banu, byagci@sts.org

**Measure Developer if different from Measure Steward:** The Society of Thoracic Surgeons

**Measure Developer Point(s) of Contact:** Yagci, Banu, byagci@sts.org

## Additional Information

**1. Provide any supplemental materials, if needed, as an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be collated one file with a table of contents or bookmarks. If material pertains to a specific criterion, that should be indicated.**

**[Response Begins]**

**[Response Ends]**

**2. List the workgroup/panel members' names and organizations.**

*Describe the members' role in measure development.*

**[Response Begins]**

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 National Databases 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; All Children's Hospital/Johns Hopkins University, Saint Petersburg, 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 Loddell, 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

### Task Force on Quality Initiatives

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

William Caine, MD; Intermountain Heart Institute, Murray, UT

Chris Feindel, MD; Cardiovascular Surgery Associates, Toronto, Ontario

Kristopher George, MD; Cardiac Surgical Associates of Fresno, Fresno, CA  
Kevin Lobdell, MD; Atrium Health, Charlotte, NC  
Edward Savage, MD; Cleveland Clinic/Martin Health, Stuart, FL

**[Response Ends]**

**3. Indicate the year the measure was first released.**

**[Response Begins]**

**[Response Ends]**

**4. Indicate the month and year of the most recent revision.**

**[Response Begins]**

**[Response Ends]**

**5. Indicate the frequency of review, or an update schedule, for this measure.**

**[Response Begins]**

annually

**[Response Ends]**

**6. Indicate the next scheduled update or review of this measure.**

**[Response Begins]**

**[Response Ends]**

**7. Provide a copyright statement, if applicable. Otherwise, indicate "N/A".**

**[Response Begins]**

**[Response Ends]**

**8. State any disclaimers, if applicable. Otherwise, indicate "N/A".**

**[Response Begins]**

**[Response Ends]**

**9. Provide any additional information or comments, if applicable. Otherwise, indicate "N/A".**

**[Response Begins]**

**[Response Ends]**