



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 #: 2561

Corresponding Measures:

De.2. Measure Title: STS Aortic Valve Replacement (AVR) Composite Score

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

De.3. Brief Description of Measure: STS AVR Composite Score comprises two domains consisting of six measures: Domain 1) Absence of Operative Mortality – Proportion of patients (risk-adjusted) who do not experience operative mortality. Operative mortality is defined as death during the same hospitalization as surgery or after discharge but within 30 days of the procedure; and Domain 2) Absence of Major Morbidity – Proportion of patients (risk-adjusted) who do not experience any major morbidity. Major morbidity is defined as having at least one of the following adverse outcomes: 1. reoperations for any cardiac reason, 2. renal failure, 3. deep sternal wound infection, 4. prolonged ventilation/intubation, and 5. cerebrovascular accident/permanent stroke. All measures are based on audited clinical data collected in a prospective registry and are risk-adjusted.

Participants receive a score for each of the two domains, plus an overall composite score. The overall composite score was created by “rolling up” the domain scores into a single number. In addition to receiving a numeric score, participants are assigned to rating categories designated by one star (below average performance), two stars (average performance), or three stars (above average performance). Star ratings are publicly reported on the STS website and are also currently reported on the Consumer Reports website.

1b.1. Developer Rationale: N/A

S.4. Numerator Statement: Due to the complex methodology used to construct the composite measure, it is impractical to separately discuss the numerator and denominator. The following discussion describes how each domain score is calculated and how these are combined into an overall composite score.

The STS AVR Composite Score comprises two domains consisting of six individual measures:

1. Absence of Operative Mortality

NQF # 0120 Risk-Adjusted Operative Mortality for AVR

2. Absence of Major Morbidity, scored any-or-none. The measures used are the same morbidity outcomes included in NQF #0696 STS CABG Composite Score.

Risk-Adjusted Postoperative Stroke/Cerebrovascular Accident

Risk-Adjusted Postoperative Surgical Re-exploration

Risk-Adjusted Postoperative Deep Sternal Wound Infection Rate

Risk-Adjusted Postoperative Renal Failure

Risk-Adjusted Postoperative Prolonged Intubation (Ventilation)

Participants receive a score for each of the two domains, plus an overall composite score. The overall composite score is created by “rolling up” the domain scores into a single number. In addition to receiving a numeric score, participants are assigned to rating categories designated by one star (below average performance), two stars (average performance), or three stars (above average performance).

Patient Population: The analysis population consists of adult patients aged 18 years or older who undergo isolated AVR surgery

Time Period: 3 years

Data Completeness Requirement: Participants are excluded from the analysis if they have fewer than 10 isolated AVR procedures in the patient population.

Technical Details

The unit of measurement for the STS AVR Composite Score can be either a participant (most often a cardiac surgical practice but occasionally an individual surgeon) or a hospital.

For the Absence of Operative Mortality domain, the NUMERATOR is:

Number of patients undergoing isolated AVR who survived until after discharge and >30 days post-surgery

For the Absence of Major Morbidity domain, the NUMERATOR is:

Number of patients undergoing isolated AVR who did not experience any of the five specified major morbidity endpoints*

*Morbidity endpoints consist of postoperative stroke/cerebrovascular accident, surgical re-exploration, deep sternal wound infection, renal failure, prolonged intubation (ventilation). Patients with documented history of renal failure (i.e., dialysis or baseline serum creatinine of 4.0 or higher) are excluded when counting renal failure outcomes.

STS AVR risk models are used to estimate expected rates of mortality and any-or-none morbidity (Reference: O'Brien SM, Shahian DM, Filardo G, et al. The Society of Thoracic Surgeons 2008 cardiac surgery risk models: part 2—isolated valve surgery. Ann Thorac Surg 2009;88(1 Suppl):S23–42). To enhance interpretation, mortality rates are converted to survival rates (risk-standardized survival rate = $100 - \text{risk-standardized mortality rate}$), and morbidity rates are converted to “absence of morbidity” rates (risk-standardized absence of morbidity rate = $100 - \text{risk-standardized morbidity rate}$). Defining scores in this manner ensures that increasingly positive values reflect better performance, which is easier for consumers to interpret.

(Please see the appendix for the formula used to calculate the overall composite score.)

The method is equivalent to calculating a weighted average, with weights proportional to the inverse of the SD. In the most recent production of the STS AVR Composite Score based on data from July 2010 – June 2013, wtmort=0.79 and wtmorb = 0.21.

Star Rating: Star ratings are derived by testing whether the participant's composite or domain score is significantly different from the overall STS average. For instance, if for each of the 2 composite score domains, a participant's estimated score is lower than the overall STS average, but the difference between the participant and STS is not statistically significant, the ratings would each be 2 stars. If however, for the overall composite, the point estimate is lower than the STS average, AND this difference is statistically significant, the overall participant star rating is 1 star. The fact that statistical significance was achieved for the composite score but not the individual domains reflects the greater precision of the composite score compared to individual endpoints. This precision is achieved by aggregating information across multiple endpoints instead of a single endpoint.

Additional details regarding the AVR Composite Score are provided in the attached manuscript:

Shahian DM, He X, Jacobs JP, et al. The Society of Thoracic Surgeons Isolated Aortic Valve Replacement (AVR) Composite Score: a report of the STS Quality Measurement Task Force. Ann Thorac Surg 2012;94:2166-71.

S.6. Denominator Statement: Due to the complex methodology used to construct the composite measure, it is impractical to separately discuss the numerator and denominator. The following discussion describes how each domain score is calculated and how these are combined into an overall composite score.

The STS AVR Composite Score comprises two domains consisting of six individual measures:

1. Absence of Operative Mortality

NQF # 0120 Risk-Adjusted Operative Mortality for AVR

2. Absence of Major Morbidity, scored any-or-none. The measures used are the same morbidity outcomes included in NQF #0696 STS CABG Composite Score.

Risk-Adjusted Postoperative Stroke/Cerebrovascular Accident

Risk-Adjusted Postoperative Surgical Re-exploration
Risk-Adjusted Postoperative Deep Sternal Wound Infection Rate
Risk-Adjusted Postoperative Renal Failure
Risk-Adjusted Postoperative Prolonged Intubation (Ventilation)

Participants receive a score for each of the two domains, plus an overall composite score. The overall composite score is created by “rolling up” the domain scores into a single number. In addition to receiving a numeric score, participants are assigned to rating categories designated by one star (below average performance), two stars (average performance), or three stars (above average performance).

Patient Population: The analysis population consists of adult patients aged 18 years or older who undergo isolated AVR surgery

Time Period: 3 years

Data Completeness Requirement: Participants are excluded from the analysis if they have fewer than 10 isolated AVR procedures in the patient population.

Technical Details

The unit of measurement for the STS AVR Composite Score can be either a participant (most often a cardiac surgical practice but occasionally an individual surgeon) or a hospital.

For the Absence of Operative Mortality domain AND the Absence of Major Morbidity domain, the DENOMINATOR is:
Number of patients undergoing isolated AVR during the measurement period

STS AVR risk models are used to estimate expected rates of mortality and any-or-none morbidity (Reference: O’Brien SM, Shahian DM, Filardo G, et al. The Society of Thoracic Surgeons 2008 cardiac surgery risk models: part 2—isolated valve surgery. Ann Thorac Surg 2009;88(1 Suppl):S23–42). To enhance interpretation, mortality rates are converted to survival rates (risk-standardized survival rate = $100 - \text{risk-standardized mortality rate}$), and morbidity rates are converted to “absence of morbidity” rates (risk-standardized absence of morbidity rate = $100 - \text{risk-standardized morbidity rate}$). Defining scores in this manner ensures that increasingly positive values reflect better performance, which is easier for consumers to interpret.

(Please see the appendix for the formula used to calculate the overall composite score.)

The method is equivalent to calculating a weighted average, with weights proportional to the inverse of the SD. In the most recent production of the STS AVR Composite Score based on data from July 2010 – June 2013, wtmort=0.79 and wtmorb = 0.21.

Star Rating: Star ratings are derived by testing whether the participant’s composite or domain score is significantly different from the overall STS average. For instance, if for each of the 2 composite score domains, a participant’s estimated score is lower than the overall STS average, but the difference between the participant and STS is not statistically significant, the ratings would each be 2 stars. If however, for the overall composite, the point estimate is lower than the STS average, AND this difference is statistically significant, the overall participant star rating is 1 star. The fact that statistical significance was achieved for the composite score but not the individual domains reflects the greater precision of the composite score compared to individual endpoints. This precision is achieved by aggregating information across multiple endpoints instead of a single endpoint.

Additional details regarding the AVR Composite Score are provided in the attached manuscript:
Shahian DM, He X, Jacobs JP, et al. The Society of Thoracic Surgeons Isolated Aortic Valve Replacement (AVR) Composite Score: a report of the STS Quality Measurement Task Force. Ann Thorac Surg 2012;94:2166-71.

S.8. Denominator Exclusions: Please see S.6 above

De.1. Measure Type: Composite

S.17. Data Source: Registry Data

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

IF Endorsement Maintenance – Original Endorsement Date: Nov 07, 2014 **Most Recent Endorsement Date:** Jun 10, 2019

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, Performance Gap, Priority – 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

[2561_Evidence_Form.STS_AVR_Composite_Score_0318-Maint-Final.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

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.

N/A

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.

[Distribution of STS isolated AVR composite measure in the latest four STS harvests for which the measure was reported](#)

| Stat | STS Harvests* | | | |
|---------------|---------------|-------------|-----------|-------------|
| | Latest | Spring 2017 | Fall 2016 | Spring 2016 |
| # Participant | 923 | 921 | 840 | 986 |
| # Operations | 81531 | 83549 | 73792 | 89392 |
| Mean | 0.955 | 0.953 | 0.953 | 0.948 |
| STD | 0.0117 | 0.013 | 0.0133 | 0.0148 |
| IQR | 0.0136 | 0.017 | 0.0164 | 0.0172 |
| Percentiles | | | | |
| 0% | 0.897 | 0.872 | 0.880 | 0.859 |
| 10% | 0.940 | 0.936 | 0.936 | 0.930 |
| 20% | 0.946 | 0.943 | 0.944 | 0.938 |
| 30% | 0.951 | 0.948 | 0.948 | 0.943 |
| 40% | 0.954 | 0.952 | 0.952 | 0.947 |
| 50% | 0.956 | 0.955 | 0.955 | 0.951 |
| 60% | 0.959 | 0.958 | 0.958 | 0.954 |
| 70% | 0.961 | 0.961 | 0.961 | 0.957 |
| 80% | 0.964 | 0.964 | 0.964 | 0.960 |

| | | | | |
|-----|-------|-------|-------|-------|
| 90% | 0.968 | 0.967 | 0.968 | 0.964 |
|-----|-------|-------|-------|-------|

| | | | | |
|------|-------|-------|-------|-------|
| 100% | 0.979 | 0.979 | 0.980 | 0.977 |
|------|-------|-------|-------|-------|

US Geographic Region

| | | | | |
|---------|-----|-----|-----|-----|
| Midwest | 258 | 258 | 244 | 270 |
|---------|-----|-----|-----|-----|

| | | | | |
|-----------|-----|-----|-----|-----|
| Northeast | 130 | 128 | 114 | 137 |
|-----------|-----|-----|-----|-----|

| | | | | |
|-------|----|---|---|---|
| Other | 10 | 6 | 4 | 1 |
|-------|----|---|---|---|

| | | | | |
|-------|-----|-----|-----|-----|
| South | 335 | 338 | 293 | 361 |
|-------|-----|-----|-----|-----|

| | | | | |
|------|-----|-----|-----|-----|
| West | 190 | 191 | 185 | 217 |
|------|-----|-----|-----|-----|

* Composite measure analysis of each harvest uses the most recent three year data until the end of last quarter. For example Spring 2013 harvest uses data until December 2012.

Changes of scores between measures calculated with data from Fall 2017 (July 2014 - June 2017) and Spring 2016 (Jan 2013 - Dec 2015)

(Please see Appendix for this scatter plot. Please also see Appendix if table of performance values does not display clearly above.)

The Spearman rank correlation of the measure between the two time periods is 0.67. The Pearson correlation is 0.68.

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.

This composite measure gauges the performance of STS participant (typically a hospital, a hospital group, or a surgeon group). It is not a patient or operation level measure. Therefore we do not provide data stratified by patient characteristics. Instead, we provide results stratified by participant characteristics.

Distribution of isolated AVR composite measures by regions, Fall 2017 harvest, July 2014 - June 2017.

| Stat | Midwest | Northeast | South | West | Other |
|------|---------|-----------|-------|------|-------|
|------|---------|-----------|-------|------|-------|

| | | | | | |
|---------------|-----|-----|-----|-----|----|
| # Participant | 258 | 130 | 335 | 190 | 10 |
|---------------|-----|-----|-----|-----|----|

| | | | | | |
|--------------|-------|-------|-------|-------|-----|
| # Operations | 20157 | 18238 | 26573 | 15672 | 891 |
|--------------|-------|-------|-------|-------|-----|

| | | | | | |
|------|-------|-------|-------|-------|-------|
| Mean | 0.955 | 0.959 | 0.953 | 0.954 | 0.952 |
|------|-------|-------|-------|-------|-------|

| | | | | | |
|-----|--------|--------|--------|--------|--------|
| STD | 0.0106 | 0.0102 | 0.0127 | 0.0114 | 0.0113 |
|-----|--------|--------|--------|--------|--------|

| | | | | | |
|-----|--------|--------|--------|--------|--------|
| IQR | 0.0119 | 0.0124 | 0.0141 | 0.0158 | 0.0104 |
|-----|--------|--------|--------|--------|--------|

Percentiles

| | | | | | |
|----|-------|-------|-------|-------|-------|
| 0% | 0.915 | 0.927 | 0.897 | 0.919 | 0.927 |
|----|-------|-------|-------|-------|-------|

| | | | | | |
|-----|-------|-------|-------|-------|-------|
| 10% | 0.941 | 0.947 | 0.939 | 0.939 | 0.942 |
|-----|-------|-------|-------|-------|-------|

| | | | | | |
|-----|-------|-------|-------|-------|-------|
| 20% | 0.947 | 0.952 | 0.946 | 0.945 | 0.944 |
|-----|-------|-------|-------|-------|-------|

| | | | | | |
|-----|-------|-------|-------|-------|-------|
| 30% | 0.951 | 0.956 | 0.949 | 0.950 | 0.950 |
|-----|-------|-------|-------|-------|-------|

| | | | | | |
|-----|-------|-------|-------|-------|-------|
| 40% | 0.954 | 0.958 | 0.953 | 0.953 | 0.953 |
|-----|-------|-------|-------|-------|-------|

| | | | | | |
|-----|-------|-------|-------|-------|-------|
| 50% | 0.957 | 0.961 | 0.955 | 0.956 | 0.954 |
|-----|-------|-------|-------|-------|-------|

| | | | | | |
|-----|-------|-------|-------|-------|-------|
| 60% | 0.958 | 0.963 | 0.958 | 0.958 | 0.955 |
|-----|-------|-------|-------|-------|-------|

| | | | | | |
|-----|-------|-------|-------|-------|-------|
| 70% | 0.961 | 0.965 | 0.960 | 0.961 | 0.956 |
|-----|-------|-------|-------|-------|-------|

| | | | | | |
|-----|-------|-------|-------|-------|-------|
| 80% | 0.963 | 0.968 | 0.963 | 0.965 | 0.957 |
|-----|-------|-------|-------|-------|-------|

| | | | | | |
|-----|-------|-------|-------|-------|-------|
| 90% | 0.966 | 0.971 | 0.967 | 0.967 | 0.959 |
|-----|-------|-------|-------|-------|-------|

| | | | | | |
|------|-------|-------|-------|-------|-------|
| 100% | 0.976 | 0.978 | 0.977 | 0.979 | 0.970 |
|------|-------|-------|-------|-------|-------|

Distribution of isolated AVR composite measures by regions, Spring 2016 harvest, Jan 2013 - Dec 2015

#2561 STS Aortic Valve Replacement (AVR) Composite Score, Last Updated: Jun 10, 2019

| Stat | Midwest | Northeast | South | West | Other |
|---------------|---------|-----------|--------|--------|-------|
| # Participant | 270 | 137 | 361 | 217 | 1 |
| # Operations | 21608 | 19590 | 29614 | 18556 | 24 |
| Mean | 0.948 | 0.954 | 0.947 | 0.949 | 0.953 |
| STD | 0.0129 | 0.0122 | 0.0169 | 0.014 | NA |
| IQR | 0.0157 | 0.0158 | 0.0189 | 0.0185 | NA |
| Percentiles | | | | | |
| 0% | 0.869 | 0.907 | 0.859 | 0.900 | NA |
| 10% | 0.932 | 0.940 | 0.927 | 0.933 | NA |
| 20% | 0.938 | 0.945 | 0.935 | 0.938 | NA |
| 30% | 0.943 | 0.949 | 0.941 | 0.943 | NA |
| 40% | 0.947 | 0.952 | 0.946 | 0.947 | NA |
| 50% | 0.950 | 0.956 | 0.950 | 0.949 | NA |
| 60% | 0.953 | 0.957 | 0.954 | 0.954 | NA |
| 70% | 0.955 | 0.961 | 0.957 | 0.958 | NA |
| 80% | 0.958 | 0.964 | 0.960 | 0.960 | NA |
| 90% | 0.961 | 0.968 | 0.964 | 0.965 | NA |
| 100% | 0.974 | 0.977 | 0.971 | 0.975 | NA |

(Please see Appendix if tables do not display clearly above.) This composite measure gauges the performance of STS participant (typically a hospital, a hospital group, or a surgeon group). It is not a patient or operation level measure. Therefore we do not provide data stratified by patient characteristics. Instead, we provide results stratified by participant characteristics.

Distribution of isolated AVR composite measures by regions, Fall 2017 harvest, July 2014 - June 2017.

| Stat | Midwest | Northeast | South | West | Other |
|---------------|---------|-----------|--------|--------|--------|
| # Participant | 258 | 130 | 335 | 190 | 10 |
| # Operations | 20157 | 18238 | 26573 | 15672 | 891 |
| Mean | 0.955 | 0.959 | 0.953 | 0.954 | 0.952 |
| STD | 0.0106 | 0.0102 | 0.0127 | 0.0114 | 0.0113 |
| IQR | 0.0119 | 0.0124 | 0.0141 | 0.0158 | 0.0104 |
| Percentiles | | | | | |
| 0% | 0.915 | 0.927 | 0.897 | 0.919 | 0.927 |
| 10% | 0.941 | 0.947 | 0.939 | 0.939 | 0.942 |
| 20% | 0.947 | 0.952 | 0.946 | 0.945 | 0.944 |
| 30% | 0.951 | 0.956 | 0.949 | 0.950 | 0.950 |
| 40% | 0.954 | 0.958 | 0.953 | 0.953 | 0.953 |
| 50% | 0.957 | 0.961 | 0.955 | 0.956 | 0.954 |
| 60% | 0.958 | 0.963 | 0.958 | 0.958 | 0.955 |
| 70% | 0.961 | 0.965 | 0.960 | 0.961 | 0.956 |
| 80% | 0.963 | 0.968 | 0.963 | 0.965 | 0.957 |
| 90% | 0.966 | 0.971 | 0.967 | 0.967 | 0.959 |
| 100% | 0.976 | 0.978 | 0.977 | 0.979 | 0.970 |

Distribution of isolated AVR composite measures by regions, Spring 2016 harvest, Jan 2013 - Dec 2015

| Stat | Midwest | Northeast | South | West | Other |
|---------------|---------|-----------|--------|--------|-------|
| # Participant | 270 | 137 | 361 | 217 | 1 |
| # Operations | 21608 | 19590 | 29614 | 18556 | 24 |
| Mean | 0.948 | 0.954 | 0.947 | 0.949 | 0.953 |
| STD | 0.0129 | 0.0122 | 0.0169 | 0.014 | NA |
| IQR | 0.0157 | 0.0158 | 0.0189 | 0.0185 | NA |
| Percentiles | | | | | |

| | | | | | |
|------|-------|-------|-------|-------|----|
| 0% | 0.869 | 0.907 | 0.859 | 0.900 | NA |
| 10% | 0.932 | 0.940 | 0.927 | 0.933 | NA |
| 20% | 0.938 | 0.945 | 0.935 | 0.938 | NA |
| 30% | 0.943 | 0.949 | 0.941 | 0.943 | NA |
| 40% | 0.947 | 0.952 | 0.946 | 0.947 | NA |
| 50% | 0.950 | 0.956 | 0.950 | 0.949 | NA |
| 60% | 0.953 | 0.957 | 0.954 | 0.954 | NA |
| 70% | 0.955 | 0.961 | 0.957 | 0.958 | NA |
| 80% | 0.958 | 0.964 | 0.960 | 0.960 | NA |
| 90% | 0.961 | 0.968 | 0.964 | 0.965 | NA |
| 100% | 0.974 | 0.977 | 0.971 | 0.975 | NA |

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

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

1c. Composite Quality Construct and Rationale

1c.1. A composite performance measure is a combination of two or more component measures, each of which individually reflects quality of care, into a single performance measure with a single score.

For purposes of NQF measure submission, evaluation, and endorsement, the following will be considered composites:

- Measures with two or more individual performance measure scores combined into one score for an accountable entity.
- Measures with two or more individual component measures assessed separately for each patient and then aggregated into one score for an accountable entity:
 - all-or-none measures (e.g., all essential care processes received, or outcomes experienced, by each patient);

1c.1. Please identify the composite measure construction: **two or more individual performance measure scores combined into one score**

1c.2. Describe the quality construct, including:

- the overall area of quality
- included component measures and
- the relationship of the component measures to the overall composite and to each other.

The STS Aortic Valve Replacement (AVR) Composite Score measures surgical performance based on a combination of the NQF-endorsed risk-adjusted operative mortality outcome measure and the same morbidity outcomes that make up the NQF-endorsed risk-adjusted morbidity measures for isolated CABG. An NQF-endorsed structure measure, database participation, is included de facto as only STS Adult Cardiac Surgery Database participants are eligible to receive composite scores. To assess overall quality, the composite comprises the following two domains:

Domain 1 – Absence of Operative Mortality

Proportion of patients (risk-adjusted) who do not experience operative mortality. Operative mortality is defined as death during the same hospitalization as surgery or after discharge but within 30 days of the procedure; and

Domain 2 – Absence of Major Morbidity

Proportion of patients (risk-adjusted) who do not experience any major morbidity. Major morbidity is defined as having at least one of the following adverse outcomes: 1) reoperations for any cardiac reason, 2) renal failure, 3) deep sternal wound infection, 4) prolonged ventilation/intubation, 5) cerebrovascular accident/permanent stroke.

Participants receive a score for each of the two domains, plus an overall composite score. The overall composite score was created by “rolling up” the domain scores into a single number. In addition to receiving a numeric score, participants are assigned to rating categories designated by one to three stars.

This composite measure differs from the NQF-endorsed STS CABG Composite Score in that it includes no process measures. This reflects the fact that, in comparison with CABG surgery, there are currently no widely accepted process measures for AVR that would meet the criteria for a performance metric. The STS Aortic Valve Replacement (AVR) Composite Score measures surgical performance based on a combination of the NQF-endorsed risk-adjusted operative mortality outcome measure and the same morbidity outcomes that make up the NQF-endorsed risk-adjusted morbidity measures for isolated CABG. An NQF-endorsed structure measure, database participation, is included de facto as only STS Adult Cardiac Surgery Database participants are eligible to receive composite scores. To assess overall quality, the composite comprises the following two domains:

Domain 1 – Absence of Operative Mortality

Proportion of patients (risk-adjusted) who do not experience operative mortality. Operative mortality is defined as death during the same hospitalization as surgery or after discharge but within 30 days of the procedure; and

Domain 2 – Absence of Major Morbidity

Proportion of patients (risk-adjusted) who do not experience any major morbidity. Major morbidity is defined as having at least one of the following adverse outcomes: 1) reoperations for any cardiac reason, 2) renal failure, 3) deep sternal wound infection, 4) prolonged ventilation/intubation, 5) cerebrovascular accident/permanent stroke.

Participants receive a score for each of the two domains, plus an overall composite score. The overall composite score was created by “rolling up” the domain scores into a single number. In addition to receiving a numeric score, participants are assigned to rating categories designated by one to three stars.

This composite measure differs from the NQF-endorsed STS CABG Composite Score in that it includes no process measures. This reflects the fact that, in comparison with CABG surgery, there are currently no widely accepted process measures for AVR that would meet the criteria for a performance metric.

1c.3. Describe the rationale for constructing a composite measure, including how the composite provides a distinctive or additive value over the component measures individually.

Risk-adjusted mortality has historically been the dominant outcomes metric for cardiac surgery procedures, but in an era when the average mortality rates for these procedures have declined to very low levels, differentiating performance based on mortality alone is inadequate. Specifically, it fails to take into account the fact that not all operative survivors receive equal quality care, and therefore enjoy an equivalent outcome, e.g., patients who survive surgery but have a debilitating complication that may substantially impact long-term quality of life, or freedom from cardiac events. This composite provides a more comprehensive measure of overall quality and is timely due to the fact that the number of isolated AVRs has nearly doubled over the past ten years, while the number of CABG operations has been relatively stable. By measuring performance for both procedures, the overall quality of a typical cardiac surgical practice is more comprehensively assessed.

1c.4. Describe how the aggregation and weighting of the component measures are consistent with the stated quality construct and rationale.

The mortality domain corresponds to a single measure, while for the morbidity domain, the study endpoint was defined in a manner that combines multiple measures and is thus a composite endpoint. To enhance interpretation, mortality rates were converted to survival rates (risk-standardized survival rate = 100 – risk-standardized mortality rate), and morbidity rates were converted to “absence of morbidity” rates (risk-standardized absence of morbidity rate = 100 – risk-standardized morbidity rate). Defining scores in this manner ensures that increasingly positive values reflect better performance.

A participant’s overall composite performance score is calculated as a weighted average of the domain-specific estimates. The weight that is applied to a given domain is inversely proportional to the standard deviation of the domain-specific scores (calculated across hospitals).

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 : Complications, Safety : Healthcare Associated Infections

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

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/documents/ACSD_DataCollectionFormV2_9_Annotated.pdf;

https://www.sts.org/sites/default/files/documents/ACSD_DataSpecificationsV2_9.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)

Attachment Attachment: S.2b._S.15._Detailed_Risk_Model_Specifications.STS_AVR_Composite_Score.docx

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 important changes to the measure specifications since last measure update and explain the reasons.

N/A

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).

Due to the complex methodology used to construct the composite measure, it is impractical to separately discuss the numerator and denominator. The following discussion describes how each domain score is calculated and how these are combined into an overall composite score.

The STS AVR Composite Score comprises two domains consisting of six individual measures:

1. Absence of Operative Mortality

NQF # 0120 Risk-Adjusted Operative Mortality for AVR

2. Absence of Major Morbidity, scored any-or-none. The measures used are the same morbidity outcomes included in NQF #0696 STS CABG Composite Score.

Risk-Adjusted Postoperative Stroke/Cerebrovascular Accident

Risk-Adjusted Postoperative Surgical Re-exploration

Risk-Adjusted Postoperative Deep Sternal Wound Infection Rate

Risk-Adjusted Postoperative Renal Failure

Risk-Adjusted Postoperative Prolonged Intubation (Ventilation)

Participants receive a score for each of the two domains, plus an overall composite score. The overall composite score is created by “rolling up” the domain scores into a single number. In addition to receiving a numeric score, participants are assigned to rating categories designated by one star (below average performance), two stars (average performance), or three stars (above average performance).

Patient Population: The analysis population consists of adult patients aged 18 years or older who undergo isolated AVR surgery

Time Period: 3 years

Data Completeness Requirement: Participants are excluded from the analysis if they have fewer than 10 isolated AVR procedures in the patient population.

Technical Details

The unit of measurement for the STS AVR Composite Score can be either a participant (most often a cardiac surgical practice but occasionally an individual surgeon) or a hospital.

For the Absence of Operative Mortality domain, the NUMERATOR is:

Number of patients undergoing isolated AVR who survived until after discharge and >30 days post-surgery

For the Absence of Major Morbidity domain, the NUMERATOR is:

Number of patients undergoing isolated AVR who did not experience any of the five specified major morbidity endpoints*

*Morbidity endpoints consist of postoperative stroke/cerebrovascular accident, surgical re-exploration, deep sternal wound infection, renal failure, prolonged intubation (ventilation). Patients with documented history of renal failure (i.e., dialysis or baseline serum creatinine of 4.0 or higher) are excluded when counting renal failure outcomes.

STS AVR risk models are used to estimate expected rates of mortality and any-or-none morbidity (Reference: O’Brien SM, Shahian DM, Filardo G, et al. The Society of Thoracic Surgeons 2008 cardiac surgery risk models: part 2—isolated valve surgery. Ann Thorac Surg 2009;88(1 Suppl):S23–42). To enhance interpretation, mortality rates are converted to survival rates (risk-standardized survival rate = $100 - \text{risk-standardized mortality rate}$), and morbidity rates are converted to “absence of morbidity” rates (risk-standardized absence of morbidity rate = $100 - \text{risk-standardized morbidity rate}$). Defining scores in this manner ensures that increasingly positive values reflect better performance, which is easier for consumers to interpret.

(Please see the appendix for the formula used to calculate the overall composite score.)

The method is equivalent to calculating a weighted average, with weights proportional to the inverse of the SD. In the most recent production of the STS AVR Composite Score based on data from July 2010 – June 2013, $\text{wtmort}=0.79$ and $\text{wtmorb} = 0.21$.

Star Rating: Star ratings are derived by testing whether the participant’s composite or domain score is significantly different from the overall STS average. For instance, if for each of the 2 composite score domains, a participant’s estimated score is lower than the overall STS average, but the difference between the participant and STS is not statistically significant, the ratings would each be 2 stars. If however, for the overall composite, the point estimate is lower than the STS average, AND this difference is statistically significant, the overall participant star rating is 1 star. The fact that statistical significance was achieved for the composite score but

not the individual domains reflects the greater precision of the composite score compared to individual endpoints. This precision is achieved by aggregating information across multiple endpoints instead of a single endpoint.

Additional details regarding the AVR Composite Score are provided in the attached manuscript: Shahian DM, He X, Jacobs JP, et al. The Society of Thoracic Surgeons Isolated Aortic Valve Replacement (AVR) Composite Score: a report of the STS Quality Measurement Task Force. *Ann Thorac Surg* 2012;94:2166-71.

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).

Please see S.4 above

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

Due to the complex methodology used to construct the composite measure, it is impractical to separately discuss the numerator and denominator. The following discussion describes how each domain score is calculated and how these are combined into an overall composite score.

The STS AVR Composite Score comprises two domains consisting of six individual measures:

1. Absence of Operative Mortality

NQF # 0120 Risk-Adjusted Operative Mortality for AVR

2. Absence of Major Morbidity, scored any-or-none. The measures used are the same morbidity outcomes included in NQF #0696 STS CABG Composite Score.

Risk-Adjusted Postoperative Stroke/Cerebrovascular Accident

Risk-Adjusted Postoperative Surgical Re-exploration

Risk-Adjusted Postoperative Deep Sternal Wound Infection Rate

Risk-Adjusted Postoperative Renal Failure

Risk-Adjusted Postoperative Prolonged Intubation (Ventilation)

Participants receive a score for each of the two domains, plus an overall composite score. The overall composite score is created by “rolling up” the domain scores into a single number. In addition to receiving a numeric score, participants are assigned to rating categories designated by one star (below average performance), two stars (average performance), or three stars (above average performance).

Patient Population: The analysis population consists of adult patients aged 18 years or older who undergo isolated AVR surgery

Time Period: 3 years

Data Completeness Requirement: Participants are excluded from the analysis if they have fewer than 10 isolated AVR procedures in the patient population.

Technical Details

The unit of measurement for the STS AVR Composite Score can be either a participant (most often a cardiac surgical practice but occasionally an individual surgeon) or a hospital.

For the Absence of Operative Mortality domain AND the Absence of Major Morbidity domain, the DENOMINATOR is:

Number of patients undergoing isolated AVR during the measurement period

STS AVR risk models are used to estimate expected rates of mortality and any-or-none morbidity (Reference: O’Brien SM, Shahian DM, Filardo G, et al. The Society of Thoracic Surgeons 2008 cardiac surgery risk models: part 2—isolated valve surgery. *Ann Thorac*

Surg 2009;88(1 Suppl):S23–42). To enhance interpretation, mortality rates are converted to survival rates (risk-standardized survival rate = $100 - \text{risk-standardized mortality rate}$), and morbidity rates are converted to “absence of morbidity” rates (risk-standardized absence of morbidity rate = $100 - \text{risk-standardized morbidity rate}$). Defining scores in this manner ensures that increasingly positive values reflect better performance, which is easier for consumers to interpret.

(Please see the appendix for the formula used to calculate the overall composite score.)

The method is equivalent to calculating a weighted average, with weights proportional to the inverse of the SD. In the most recent production of the STS AVR Composite Score based on data from July 2010 – June 2013, $\text{wtmort} = 0.79$ and $\text{wtmorb} = 0.21$.

Star Rating: Star ratings are derived by testing whether the participant’s composite or domain score is significantly different from the overall STS average. For instance, if for each of the 2 composite score domains, a participant’s estimated score is lower than the overall STS average, but the difference between the participant and STS is not statistically significant, the ratings would each be 2 stars. If however, for the overall composite, the point estimate is lower than the STS average, AND this difference is statistically significant, the overall participant star rating is 1 star. The fact that statistical significance was achieved for the composite score but not the individual domains reflects the greater precision of the composite score compared to individual endpoints. This precision is achieved by aggregating information across multiple endpoints instead of a single endpoint.

Additional details regarding the AVR Composite Score are provided in the attached manuscript:

Shahian DM, He X, Jacobs JP, et al. The Society of Thoracic Surgeons Isolated Aortic Valve Replacement (AVR) Composite Score: a report of the STS Quality Measurement Task Force. Ann Thorac Surg 2012;94:2166-71.

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).

Please see S.6 above

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

Please see S.6 above

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.)

Please see S.6 above

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)

Statistical risk model

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 see S.4 and S.6 above

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 2.81 (effective July 1, 2014); Version 2.9 (effective July 1, 2017)

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.*)

Please see composite calculation formula in section S.4-S.11 in the Appendix

2. Validity – See attached Measure Testing Submission Form

RESUB-2561_comp_testing_attmt__7-1_v3-0818.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.

Yes

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.

Yes

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

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).

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.

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 3 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 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.

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) |
|-----------------------|--|
| | <p>Public Reporting</p> <p>STS Public Reporting Online https://publicreporting.sts.org/ STS Public Reporting Online https://publicreporting.sts.org/</p> <p>Quality Improvement (Internal to the specific organization)</p> <p>STS Adult Cardiac Surgery Database Participants http://www.sts.org/sts-national-database/database-managers/adult-cardiac-surgery-database</p> |

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 public reporting – approximately 67% of STS Adult Cardiac Surgery Database participants are enrolled as of November 2018.

The isolated AVR composite is one of the three ACSD composite measures publicly reported on:
<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 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.

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. 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.

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 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.

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/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 table below displays the percentage distribution of star ratings for the AVR and AVR+CABG composite measures among all participants in the Adult Cardiac Surgery Database, since the AVR composite was introduced in 2011. A one-star rating indicates that a participant's composite score is significantly lower than the STS average; a 2-star rating indicates that the composite score is not statistically different from the average; a 3-star rating indicates that the composite score is significantly higher than the average.

The general trends seen in these data over time -- with more participants in the 2-star category and fewer in the 1-star category -- demonstrate some progress on improvement and reflect some degree of success in STS objectives to ensure that all cardiothoracic surgery patients receive high-quality care.

| Comp. | Stars | 2016 | 2015 | 2014 | 2013 | 2012 | 2011 |
|-------|-------|------|-------|-------|-------|------|------|
| AVR | * | 2.17 | 3.11 | 4.22 | 3.35 | 3.0 | 3.5 |
| | ** | 90.3 | 88.75 | 87.89 | 88.98 | 91.0 | 90.6 |
| | *** | 7.53 | 8.15 | 7.89 | 7.67 | 6.0 | 5.9 |

| | | | | | | | |
|-----------|-----|-------|-------|-------|------|----|----|
| AVR+ CABG | * | 2.06 | 2.49 | 2.51 | 3.14 | NA | NA |
| | ** | 92.26 | 90.72 | 90.42 | 90.7 | NA | NA |
| | *** | 5.68 | 6.79 | 7.07 | 6.17 | NA | NA |

The table below displays the percentage distribution of star ratings for the AVR and AVR+CABG composite measures among all participants in the Adult Cardiac Surgery Database, since the AVR composite was introduced in 2011. A one-star rating indicates that a participant's composite score is significantly lower than the STS average; a 2-star rating indicates that the composite score is not statistically different from the average; a 3-star rating indicates that the composite score is significantly higher than the average.

The general trends seen in these data over time -- with more participants in the 2-star category and fewer in the 1-star category -- demonstrate some progress on improvement and reflect some degree of success in STS objectives to ensure that all cardiothoracic surgery patients receive high-quality care.

| Comp. | Stars | 2016 | 2015 | 2014 | 2013 | 2012 | 2011 |
|-------|-------|------|-------|-------|-------|------|------|
| AVR | * | 2.17 | 3.11 | 4.22 | 3.35 | 3.0 | 3.5 |
| | ** | 90.3 | 88.75 | 87.89 | 88.98 | 91.0 | 90.6 |
| | *** | 7.53 | 8.15 | 7.89 | 7.67 | 6.0 | 5.9 |

| | | | | | | | |
|-----------|-----|-------|-------|-------|------|----|----|
| AVR+ CABG | * | 2.06 | 2.49 | 2.51 | 3.14 | NA | NA |
| | ** | 92.26 | 90.72 | 90.42 | 90.7 | NA | NA |
| | *** | 5.68 | 6.79 | 7.07 | 6.17 | NA | NA |

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 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.

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

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

0120 : Risk-Adjusted Operative Mortality for Aortic Valve Replacement (AVR)

0129 : Risk-Adjusted Postoperative Prolonged Intubation (Ventilation)

0130 : Risk-Adjusted Deep Sternal Wound Infection

0131 : Risk-Adjusted Stroke/Cerebrovascular Accident

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

N/A

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.

N/A

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: AVR-Comp-2561-REVISED-Appendix-Fall2018.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 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 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

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

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 Year the measure was first released: 2011

Ad.3 Month and Year of most recent revision: 07, 2017

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? 2019

Ad.6 Copyright statement:

Ad.7 Disclaimers:

Ad.8 Additional Information/Comments: