



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 #: 0119

Corresponding Measures:

Measure Title: Risk-Adjusted Operative Mortality for CABG

Measure Steward: The Society of Thoracic Surgeons

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

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. The published literature on coronary bypass grafting describe examples of services/care processes that impact operative mortality. Preoperative patient selecting, surgical timing post coronary event, intraoperative conduct of the case, and many aspects to postoperative care have all been shown to have significant impact on the operative mortality over the last few decades.

1. Ferguson TB, Hammill BG, et al. A decade of change—risk profiles and outcomes for isolated coronary artery bypass grafting procedures, 1990-1999: a report from the STS National Database Committee and the Duke Clinical Research Institute. *Ann Thorac Surg.* 2002;73(2):480-489; discussion 489-490.
2. Grover FL, Shroyer AL, et al. A decade's experience with quality improvement in cardiac surgery using the Veterans Affairs and Society of Thoracic Surgery national databases. *Ann Thorac Surg.* 2001; 234(4):464-472; discussion 472-474.
3. Hogue CW, Barzilai B, et al. Gender differences in neurologic outcomes and mortality after cardiac surgery: an STS National Database report. *Circulation.* 103:2133-2137.
4. Shroyer AL, Coombs LP, Peterson ED, et al. The Society of Thoracic Surgeons: 30-day operative mortality and morbidity risk models. *Ann Thorac Surg.* 2003;75:1856-1865.
5. Williams ML, Muhlbaier LH, Schroder JN, et. al. Risk-adjusted short- and long-term outcomes for on-pump versus off-pump coronary artery bypass surgery. *Circulation.* 2005 Aug 30;112(9 Suppl):I366-70.
6. Shroyer AL, Grover FL, Hattler B, et. al. On-pump versus off-pump coronary artery bypass surgery. *N Engl J Med.* 2009 Nov 5;361(19):1827-37.
7. Hannan EL, Wu C, Smith CR, et. al. Off-pump versus on-pump coronary artery bypass graft surgery: differences in short-term outcomes and in long-term mortality and need for subsequent revascularization. *Circulation.* 2007 Sep

4;116(10):1145-52. Epub 2007 Aug 20.

8. ElBardissi AW, Aranki SF, Sheng S, et al. Trends in isolated coronary artery bypass grafting: an analysis of the Society of Thoracic Surgeons adult cardiac surgery database. J Thorac Cardiovasc Surg. 2012 Feb;143(2):273-81.

9. Rangrass G, Ghaferi AA, Dimick. Explaining Racial Disparities in Outcomes After Cardiac Surgery: The Role of Hospital Quality. JAMA Surg. 2014;149(3):223-227.

sp.12. Numerator Statement: Number of patients undergoing isolated CABG 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 CABG

sp.16. Denominator Exclusions: N/A

Measure Type: Outcome

sp.28. Data Source:

Registry Data

sp.07. Level of Analysis:

Clinician: Group/Practice

Facility

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

Most Recent Endorsement Date: 6/10/2019 12:46:43 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. The published literature on coronary bypass grafting describe examples of services/care processes that impact operative mortality. Preoperative patient selecting, surgical timing post coronary event, intraoperative conduct of the case, and many aspects to postoperative care have all been shown to have significant impact on the operative mortality over the last few decades.

1. Ferguson TB, Hammill BG, et al. A decade of change—risk profiles and outcomes for isolated coronary artery bypass grafting procedures, 1990-1999: a report from the STS National Database Committee and the Duke Clinical Research Institute. *Ann Thorac Surg.* 2002;73(2):480-489; discussion 489-490.
2. Grover FL, Shroyer AL, et al. A decade's experience with quality improvement in cardiac surgery using the Veterans Affairs and Society of Thoracic Surgery national databases. *Ann Thorac Surg.* 2001; 234(4):464-472; discussion 472-474.
3. Hogue CW, Barzilai B, et al. Gender differences in neurologic outcomes and mortality after cardiac surgery: an STS National Database report. *Circulation.* 103:2133-2137.
4. Shroyer AL, Coombs LP, Peterson ED, et al. The Society of Thoracic Surgeons: 30-day operative mortality and morbidity risk models. *Ann Thorac Surg.* 2003;75:1856-1865.
5. Williams ML, Muhlbaier LH, Schroder JN, et. al. Risk-adjusted short- and long-term outcomes for on-pump versus off-pump coronary artery bypass surgery. *Circulation.* 2005 Aug 30;112(9 Suppl):I366-70.
6. Shroyer AL, Grover FL, Hattler B, et. al. On-pump versus off-pump coronary artery bypass surgery. *N Engl J Med.* 2009 Nov 5;361(19):1827-37.
7. Hannan EL, Wu C, Smith CR, et. al. Off-pump versus on-pump coronary artery bypass graft surgery: differences in short-term outcomes and in long-term mortality and need for subsequent revascularization. *Circulation.* 2007 Sep 4;116(10):1145-52. Epub 2007 Aug 20.
8. ElBardissi AW, Aranki SF, Sheng S, et al. Trends in isolated coronary artery bypass grafting: an analysis of the Society of Thoracic Surgeons adult cardiac surgery database. *J Thorac Cardiovasc Surg.* 2012 Feb;143(2):273-81.
9. Rangrass G, Ghaferi AA, Dimick. Explaining Racial Disparities in Outcomes After Cardiac Surgery: The Role of Hospital Quality. *JAMA Surg.* 2014;149(3):223-227.

[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 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 2016 - June 2017 and July 2015 - June 2016

Distribution July 2015 - June 2016

Odds ratio July 2015 - June 2016

Risk adjusted Rate, % July 2016 - June 2017

Odds ratio July 2016 - June 2017

Risk adjusted Rate, %

Participant 1085 1085 1072 1072

Operations 159246 159246 157700 157700

Mean 1.03 2.24 1.03 2.34

STD 0.27 0.53 0.29 0.59

IQR 0.29 0.58 0.32 0.67

0% 0.46 1.10 0.44 1.11

10% 0.76 1.71 0.75 1.75

20% 0.83 1.84 0.82 1.89

30% 0.88 1.95 0.88 2.01

40% 0.92 2.02 0.92 2.12

50% 0.98 2.14 0.97 2.21

60% 1.03 2.24 1.03 2.34

70% 1.10 2.39 1.12 2.53

80% 1.19 2.57 1.22 2.72

90% 1.35 2.90 1.37 3.05

100% 2.85 5.54 2.84 6.11

Midwest 301 301 296 296

Northeast 138 138 136 136

Other Region 13 13 18 18

South 413 413 407 407

West 220 220 215 215

*Other region = outside of the four U.S. geographic regions.

We report the measure for all STS participants with eligible cases even if the number of cases (denominator) is small. It is known that with or without risk adjustment, small number of cases yield less reliable estimates. Therefore, to facilitate a more reliable comparison across time periods, we also provided the measure summary in only participants with 50 or more cases in all 12 months of the corresponding year.

Distribution of participant-specific risk adjusted odds ratio and event rates in July 2016 - June 2017 and July 2015 - June 2016, in STS database participants who reported in each of the 12 months and a total of at least 50 eligible

cases (denominator) in the corresponding year of each

Distribution July 2015 - June 2016

Odds ratio July 2015 - June 2016

Risk adjusted Rate, % July 2016 - June 2017

Odds ratio July 2016 - June 2017

Risk adjusted Rate, %

Participant 889 889 848 848

Operations 150036 150036 147091 147091

Mean 1.02 2.23 1.03 2.32

STD 0.28 0.55 0.31 0.63

IQR 0.31 0.62 0.35 0.73

0% 0.46 1.10 0.44 1.11

10% 0.75 1.69 0.74 1.71

20% 0.81 1.81 0.80 1.85

30% 0.86 1.91 0.85 1.96

40% 0.90 1.99 0.91 2.08

50% 0.96 2.11 0.97 2.20

60% 1.02 2.24 1.02 2.32

70% 1.10 2.40 1.11 2.51

80% 1.20 2.58 1.23 2.74

90% 1.38 2.92 1.40 3.09

100% 2.85 5.54 2.84 6.11

Midwest 229 229 220 220

Northeast 125 125 118 118

Other Region 4 4 4 4

South 354 354 339 339

West 177 177 167 167

*Other region = outside of the four U.S. geographic regions.

The figures below show the changes in participant specific odds ratios and risk adjusted rates between two most recent adjust years. Only participants that reported data to STS in both years are included.

Changes of measure from July 2015 - June 2016 to July 2016 - June 2017

(Please see Appendix for these scatter plots. Please also see Appendix if tables of performance values do not display clearly above.)

The Spearman rank correlation of the measure between the two time periods is 0.34. The Pearson correlation is 0.39.

Similarly, we created the figures comparing odds ratios and risk adjusted rates from the two years in participants with more than 50 cases or more and reported data every month in each year.

Changes of measure from July 2015 - June 2016 to July 2016 - June 2017 in STS database participants who reported in each of the 12 months and a total of at least 50 eligible cases (denominator) in each year
(Please see Appendix for these scatter plots.)

The Spearman rank correlation of the measure between the two time periods in these relatively larger participants is 0.36. The Pearson correlation is 0.42.

[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 interests 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 2016 - June 2017 and July 2015 - June 2016, by sex

Distribution Male July 11 - June 12 Male July 12 - June 13 Female July 11 - June 12 Female July 12 - June 13

Participant 1085 1071 1079 1065

Operations 119746 119503 39500 38197

Mean 1.02 1.03 1.02 1.01

STD 0.24 0.27 0.21 0.14

IQR 0.26 0.28 0.22 0.16

0% 0.54 0.44 0.49 0.68

10% 0.78 0.78 0.82 0.87

20% 0.85 0.84 0.88 0.90

30% 0.89 0.89 0.91 0.93

40% 0.93 0.93 0.94 0.95

50% 0.97 0.97 0.96 0.98

60% 1.02 1.03 1.00 1.00

70% 1.09 1.10 1.08 1.06

80% 1.17 1.19 1.16 1.10

90% 1.32 1.34 1.29 1.21

100% 2.74 2.61 2.44 1.84

Distribution of participant-specific risk adjusted event rates (%) in July 2016 - June 2017 and July 2015 - June 2016, by sex

Distribution Male July 11 - June 12 Male July 12 - June 13 Female July 11 - June 12 Female July 12 - June 13

Participant 1085 1071 1079 1065

Operations 119746 119503 39500 38197

Mean 1.91 2.00 3.18 3.30

STD 0.41 0.47 0.58 0.42

IQR 0.45 0.51 0.64 0.48

0% 1.07 0.92 1.76 2.33
 10% 1.51 1.54 2.61 2.88
 20% 1.61 1.66 2.77 2.98
 30% 1.69 1.75 2.86 3.07
 40% 1.75 1.83 2.95 3.14
 50% 1.82 1.90 3.03 3.22
 60% 1.90 2.00 3.13 3.29
 70% 2.03 2.12 3.35 3.44
 80% 2.16 2.29 3.57 3.57
 90% 2.43 2.55 3.90 3.88
 100% 4.61 4.83 6.98 5.50

Distribution of participant-specific risk adjusted odds ratio in July 2016 - June 2017 and July 2015 - June 2016, by race

Distribution White July 11 - June 12 White July 12 - June 13 Black July 11 - June 12 Black July 12 - June 13 Other July 11 - June 12 Other July 12 - June 13

Participant 1078 1062 906 880 993 980

Operations 131096 129576 11923 11499 16227 16625

Mean 1.03 1.03 1.01 1.00 1.00 1.01

STD 0.25 0.27 0.13 0.0059 0.046 0.14

IQR 0.25 0.29 0.04 0.0022 0.016 0.05

0% 0.48 0.45 0.65 0.98 0.78 0.58
 10% 0.79 0.77 0.91 1.00 0.96 0.89
 20% 0.85 0.84 0.95 1.00 0.98 0.94
 30% 0.89 0.88 0.96 1.00 0.99 0.96
 40% 0.94 0.93 0.98 1.00 0.99 0.97
 50% 0.97 0.97 0.99 1.00 0.99 0.98
 60% 1.02 1.03 0.99 1.00 1.00 0.99
 70% 1.08 1.10 1.00 1.00 1.00 1.00
 80% 1.18 1.20 1.00 1.00 1.00 1.04
 90% 1.32 1.36 1.17 1.01 1.07 1.19
 100% 2.70 2.44 2.01 1.04 1.32 2.31

Distribution of participant-specific risk adjusted event rates (%) in July 2016 - June 2017 and July 2015 - June 2016, by race

Distribution White July 11 - June 12 White July 12 - June 13 Black July 11 - June 12 Black July 12 - June 13 Other July 11 - June 12 Other July 12 - June 13

Participant 1078 1062 906 880 993 980

Operations 131096 129576 11923 11499 16227 16625

Mean 2.21 2.26 2.53 2.88 2.07 2.37

STD 0.48 0.53 0.30 0.016 0.087 0.30

IQR 0.50 0.59 0.096 0.0061 0.031 0.11

0% 1.08 1.02 1.71 2.83 1.65 1.41
 10% 1.74 1.73 2.30 2.87 2.00 2.13
 20% 1.87 1.87 2.40 2.87 2.03 2.21
 30% 1.95 1.96 2.43 2.88 2.04 2.25
 40% 2.04 2.06 2.47 2.88 2.05 2.28
 50% 2.11 2.15 2.49 2.88 2.06 2.31
 60% 2.21 2.26 2.50 2.88 2.07 2.33
 70% 2.33 2.41 2.51 2.88 2.07 2.34
 80% 2.51 2.62 2.52 2.88 2.07 2.45
 90% 2.78 2.92 2.92 2.90 2.20 2.77
 100% 5.23 4.75 4.88 3.00 2.68 4.91

Distribution of participant-specific risk adjusted odds ratio in July 2016 - June 2017 and July 2015 - June 2016, by ethnicity

Distribution Hispanic ethnicity July 11 - June 12 Hispanic ethnicity July 12 - June 13 Non-Hispanic ethnicity July 11 - June 12 Non-Hispanic ethnicity July 12 - June 13

Participant 895 894 1085 1071

Operations 11803 11603 147443 146097

Mean 1.00 1.01 1.03 1.03

STD 0.059 0.15 0.27 0.27

IQR 0.017 0.043 0.28 0.30

0% 0.82 0.58 0.48 0.49

10% 0.96 0.90 0.77 0.77

20% 0.97 0.94 0.83 0.83

30% 0.99 0.96 0.88 0.88

40% 0.99 0.98 0.92 0.93

50% 0.99 0.99 0.98 0.98

60% 1.00 0.99 1.03 1.03

70% 1.00 1.00 1.10 1.11

80% 1.00 1.00 1.18 1.20

90% 1.08 1.20 1.35 1.35

100% 1.43 2.79 2.67 2.82

Distribution of participant-specific risk adjusted event rates (%) in July 2016 - June 2017 and July 2015 - June 2016, by ethnicity

Distribution Hispanic ethnicity July 11 - June 12 Hispanic ethnicity July 12 - June 13 Non-Hispanic ethnicity July 11 - June 12 Non-Hispanic ethnicity July 12 - June 13

Participant 895 894 1085 1071

Operations 11803 11603 147443 146097

Mean 2.18 2.74 2.24 2.29

STD 0.12 0.36 0.52 0.55

IQR 0.035 0.11 0.56 0.62

0% 1.84 1.68 1.11 1.16

10% 2.09 2.47 1.71 1.75

20% 2.12 2.58 1.84 1.88

30% 2.15 2.63 1.95 1.98

40% 2.16 2.66 2.03 2.09

50% 2.16 2.69 2.14 2.19

60% 2.17 2.70 2.24 2.30

70% 2.17 2.71 2.39 2.46

80% 2.17 2.72 2.55 2.65

90% 2.34 3.20 2.86 2.95

100% 3.00 5.85 5.38 5.96

(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]

Yes

[Yes Please Explain]

Denominator time frame changed to 36 months from 12 months.

[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]

The STS has developed a methodological update to our original CABG composite measure, moving it from 1 to 3-year analytic windows. This modification was done to address the sample size concerns with the original approach in contemporary practice and to ensure consistency across all STS adult cardiac composite measures. The 3-year analytic window was applied to the component measures of the CABG composite in order to be consistent and to eliminate mismatches across various sections of the feedback reports we provide to our database participants.

[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 CABG

[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 isolated CABG who die, including both 1) all deaths occurring during the hospitalization in which the CABG was performed, even if after 30 days, and 2) those deaths occurring after discharge from the hospital, but within 30 days of the procedure

[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/STSAultCVDDataCollectionFormV4_20_2_ChangesHighlightedGOLDEN12212020Annotated.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 undergoing isolated CABG 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 CABG 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), Mortality Discharge Status (MtDCStat is Dead in version 2,81 or DischMortStat is Died in Hospital in version 2.9)

Version 4.20.2

Number of isolated CABG 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 CABG

[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 CABG 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 measures 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, 4.2

[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]

01.01.2019 - 12.31.2021

[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]

Testing was based on STS data from the 36 months from January 2019 to December 2021. The testing cohort included records from 437444 isolated CABG operations performed by 1028 STS participants in North America. An STS Database participant is either a practice group of cardiothoracic surgeons or an individual cardiothoracic surgeon. **It is STS's understanding that its Adult Cardiac Surgery Database participants represent over 90% of cardiac surgery programs in the US; STS is currently in the process of determining the exact percentage.** "Isolated CABG" is defined per the STS procedure table. Frequency of included participants by geographic region is summarized in **1b**.

Distribution of participant sample sizes (denominator), and event rates (numerator/denominator)

Stat	N	% operative mortality
N	1028.0	1028.0
Mean	425.5	2.8
STD	342.3	2.3
IQR	366.8	2.0
0%	1.0	0.0
10%	113.7	1.0
20%	171.4	1.4
30%	221.0	1.7
40%	272.8	2.0
50%	331.5	2.4
60%	406.2	2.7
70%	498.8	3.2
80%	632.2	3.8
90%	838.5	4.9
100%	2508.0	40.0
*	*	*
Region		
Midwest	269	-
Northeast	136	-
Other	3	-
South	391	-
West	229	-

regions.

Other region = outside of the four U.S. geographic

* indicates that the cell is left intentionally blank

[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]

All eligible operations were included.

*	Effects	Overall N=437444
Age (years)	Median (IQR)	67.0 (59.0, 73.0)
*	Missing	0 (0.0%)
Sex	Male	334,945 (76.6%)
*	Female	102,397 (23.4%)
*	Missing	102 (0.0%)
Race	White	330,431 (75.5%)
*	Black	31,343 (7.2%)
*	Hispanic	1,476 (0.3%)
*	Asian	15,076 (3.4%)
*	Native American	2,298 (0.5%)
*	Pacific Islander	1,306 (0.3%)
*	Other	7,072 (1.6%)
*	Multiple	36,234 (8.3%)
*	Missing	12,208 (2.8%)
Hispanic or Latino Ethnicity	No	380,678 (87.0%)
*	Yes	34,720 (7.9%)
*	Missing	22,046 (5.0%)
Insurance: Younger than 65	Medicare/Medicaid	52,127 (28.6%)
*	Commercial/HMO	107,836 (59.2%)
*	None/Self Paid	13,829 (7.6%)
*	Other	8,265 (4.5%)
Insurance: 65 or Older	Medicare+Medicaid	16,696 (6.5%)
*	Medicare+Commercial without Medicaid	108,422 (42.5%)
*	Medicare without Medicaid/Commercial	130,269 (51.0%)
Region	Northeast	71,101 (16.3%)
*	South	188,438 (43.1%)
*	Midwest	98,429 (22.5%)
*	West	76,626 (17.5%)
*	Other	2,850 (0.7%)
Body Surface Area (m)	<1.5	5,135 (1.2%)
*	>=1.5 and <1.75	50,880 (11.6%)
*	>=1.75 and <2	148,882 (34.0%)
*	>=2	232,305 (53.1%)
*	Missing	242 (0.1%)
Body Mass Index (kg/m)	<18.5	2,717 (0.6%)

*	Effects	Overall N=437444
*	>=18.5 and <25	78,543 (18.0%)
*	>=25 and <30	160,856 (36.8%)
*	>=30 and <35	117,584 (26.9%)
*	>=35	77,502 (17.7%)
*	Missing	242 (0.1%)
Hematocrit	Median (IQR)	40.4 (36.6, 43.8)
*	Missing	864 (0.2%)
White Blood Cells	Median (IQR)	7.6 (6.3, 9.2)
*	Missing	1,100 (0.3%)
Platelets	Median (IQR)	7.6 (6.3, 9.2)
*	Missing	1,100 (0.3%)
Diabetes	No Diabetes	217,397 (49.7%)
*	Diabetes - Noninsulin	132,745 (30.3%)
*	Diabetes - Insulin	79,606 (18.2%)
*	Diabetes - Other	466 (0.1%)
*	Diabetes - Missing Treatment	1,889 (0.4%)
*	Diabetes - Other	4,630 (1.1%)
*	Missing	711 (0.2%)
Hypertension	No	41,644 (9.5%)
*	Yes	394,936 (90.3%)
*	Missing	864 (0.2%)
Renal Function	Creatinine <1 mg/dL	215,295 (49.2%)
*	Creatinine 1-1.5 mg/dL	172,293 (39.4%)
*	Creatinine 1.5-2 mg/dL	25,152 (5.7%)
*	Creatinine 2-2.5 mg/dL	5,461 (1.2%)
*	Creatinine >2.5 mg/dL	4,337 (1.0%)
*	Dialysis	14,060 (3.2%)
*	Missing	846 (0.2%)
Chronic Lung Disease (CLD)	None	311,365 (71.2%)
*	Mild	51,665 (11.8%)
*	Moderate	20,885 (4.8%)
*	Severe	18,791 (4.3%)
*	Severity Unknown	25,378 (5.8%)
*	Missing	9,360 (2.1%)
Peripheral Vascular Disease (PVD)	No	378,847 (86.6%)
*	Yes	57,006 (13.0%)

*	Effects	Overall N=437444
*	Missing	1,591 (0.4%)
Cerebrovascular Disease (CVD)	No	334,630 (76.5%)
*	Yes	100,008 (22.9%)
*	Missing	2,806 (0.6%)
Cerebrovascular Accident (CVA)	No CVA	398,502 (91.1%)
*	Remote CVA (> 30 days)	34,023 (7.8%)
*	Recent CVA (<= 30 days)	1,566 (0.4%)
*	CVA - Missing Timing	58 (0.0%)
*	Missing	3,295 (0.8%)
Previous Carotid Surgery	No	420,472 (96.1%)
*	Yes	14,046 (3.2%)
*	Missing	2,926 (0.7%)
Carotid Stenosis - Single	No	420,483 (96.1%)
*	Yes	8,068 (1.8%)
*	Missing	8,893 (2.0%)
Carotid Stenosis - Double	No	427,679 (97.8%)
*	Yes	872 (0.2%)
*	Missing	8,893 (2.0%)
Endocarditis	No Endocarditis	436,458 (99.8%)
*	Treated Endocarditis	297 (0.1%)
*	Active Endocarditis	30 (0.0%)
*	Endocarditis - Missing Type	4 (0.0%)
*	Missing	655 (0.1%)
Immunotherapy	No	421,184 (96.3%)
*	Yes	15,146 (3.5%)
*	Missing	1,114 (0.3%)
Pneumonia	No	400,192 (91.5%)
*	Recent	10,834 (2.5%)
*	Remote	20,228 (4.6%)
*	Missing	6,190 (1.4%)
Mediastinal Radiation	No	431,630 (98.7%)
*	Yes	3,577 (0.8%)
*	Missing	2,237 (0.5%)
Cancer within 5 years	No	409,757 (93.7%)
*	Yes	21,151 (4.8%)
*	Missing	6,536 (1.5%)

*	Effects	Overall N=437444
IV Drug Abuse	No	416,679 (95.3%)
*	Yes	14,542 (3.3%)
*	Missing	6,223 (1.4%)
Family History CAD	No	328,074 (75.0%)
*	Yes	79,469 (18.2%)
*	Missing	29,901 (6.8%)
Home Oxygen	No	430,247 (98.4%)
*	Yes	6,125 (1.4%)
*	Missing	1,072 (0.2%)
Sleep Apnea	No	353,526 (80.8%)
*	Yes	77,131 (17.6%)
*	Missing	6,787 (1.6%)
Liver Disease	No	423,473 (96.8%)
*	Yes	12,044 (2.8%)
*	Missing	1,927 (0.4%)
Unresponsive Status	No	434,799 (99.4%)
*	Yes	825 (0.2%)
*	Missing	1,820 (0.4%)
Syncope	No	421,213 (96.3%)
*	Yes	15,147 (3.5%)
*	Missing	1,084 (0.2%)
Alcohol Drinking	<= 1 drink/week	332,964 (76.1%)
*	2 - 7 drinks/week	61,472 (14.1%)
*	>= 8 drinks/week	31,676 (7.2%)
*	Missing	11,332 (2.6%)
Smoking	Never-smoker	173,163 (39.6%)
*	Past Smoker	168,532 (38.5%)
*	Current Smoker	93,975 (21.5%)
*	Missing	1,774 (0.4%)
Medications: Steroids	No	428,713 (98.0%)
*	Yes	7,624 (1.7%)
*	Missing	1,107 (0.3%)
Medications: Glycoprotein IIb-IIIa	No	432,605 (98.9%)
*	Yes	4,084 (0.9%)
*	Missing	755 (0.2%)
Medications: ADP Inhibitor	No	394,295 (90.1%)

*	Effects	Overall N=437444
*	Yes	40,424 (9.2%)
*	Missing	2,725 (0.6%)
ADP Discontinuation prior to Surgery (days)	0	6,408 (15.9%)
*	1	5,756 (14.2%)
*	2	6,450 (16.0%)
*	3	7,612 (18.8%)
*	4	8,288 (20.5%)
*	5	5,688 (14.1%)
*	Missing	222 (0.5%)
Medications: Inotropes	No	432,168 (98.8%)
*	Yes	4,903 (1.1%)
*	Missing	373 (0.1%)
Medications: ACE/ARB for non-elective status	No	342,706 (78.3%)
*	Yes	87,600 (20.0%)
*	Missing	7,138 (1.6%)
Acuity Status	Elective	164,329 (37.6%)
*	Urgent	258,082 (59.0%)
*	Emergent	14,236 (3.3%)
*	Emergent Salvage	730 (0.2%)
*	Missing	67 (0.0%)
Myocardial Infarction	No Prior MI	201,684 (46.1%)
*	MI >21 days	74,502 (17.0%)
*	MI 8-21 days	24,150 (5.5%)
*	MI 1-7 days	122,223 (27.9%)
*	MI 6-24 hrs	8,052 (1.8%)
*	MI <= 6 hrs	4,161 (1.0%)
*	MI - Missing Timing	122 (0.0%)
*	Missing	2,550 (0.6%)
Cardiac Presentation	No Symptoms	26,299 (6.0%)
*	Symptoms Unlikely to be Ischemia	48,699 (11.1%)
*	Stable Angina	71,089 (16.3%)
*	Unstable Angina	139,637 (31.9%)
*	Non-STEMI	128,776 (29.4%)
*	STEMI	22,653 (5.2%)
*	Missing	291 (0.1%)
Cardiogenic Shock	No	430,057 (98.3%)

*	Effects	Overall N=437444
*	Yes	7,050 (1.6%)
*	Missing	337 (0.1%)
Preop IABP	No	408,101 (93.3%)
*	Yes	28,296 (6.5%)
*	Missing	1,047 (0.2%)
Arrhythmia	No Arrhythmia	371,078 (84.8%)
*	Atrial Fibrillation/Flutter	42,960 (9.8%)
*	Heart Block	3,458 (0.8%)
*	Sustained VT/VF	10,245 (2.3%)
*	Multiple Types	5,224 (1.2%)
*	Arrhythmia - None of Above Types	4,048 (0.9%)
*	Arrhythmia - Missing Type	21 (0.0%)
*	Missing	410 (0.1%)
Congestive Heart Failure	No CHF	366,623 (83.8%)
*	CHF NYHA-I	3,151 (0.7%)
*	CHF NYHA-II	10,253 (2.3%)
*	CHF NYHA-III	15,329 (3.5%)
*	CHF NYHA-IV	8,479 (1.9%)
*	CHF Missing NYHA	28,816 (6.6%)
*	Missing	4,793 (1.1%)
Number of Diseased Coronary Vessels	None	3,025 (0.7%)
*	One	16,082 (3.7%)
*	Two	81,757 (18.7%)
*	Three	336,314 (76.9%)
*	Missing	266 (0.1%)
Left Main Disease >= 50%	No	296,207 (67.7%)
*	Yes	140,224 (32.1%)
*	Missing	1,013 (0.2%)
Proximal LAD Disease >= 70%	No	136,829 (31.3%)
*	Yes	299,471 (68.5%)
*	Missing	1,144 (0.3%)
Ejection Fraction (%)	Median (IQR)	55.0 (45.0, 60.0)
*	Missing	8,369 (1.9%)
Aortic Stenosis	No	413,002 (94.4%)
*	Yes	18,992 (4.3%)
*	Missing	5,450 (1.2%)

*	Effects	Overall N=437444
Mitral Stenosis	No	428,458 (97.9%)
*	Yes	3,383 (0.8%)
*	Missing	5,603 (1.3%)
Tricuspid Stenosis	No	430,324 (98.4%)
*	Yes	513 (0.1%)
*	Missing	6,607 (1.5%)
Pulmonic Stenosis	No	426,592 (97.5%)
*	Yes	305 (0.1%)
*	Missing	10,547 (2.4%)
Aortic Insufficiency	None	307,520 (70.3%)
*	Trivial	54,517 (12.5%)
*	Mild	39,790 (9.1%)
*	Moderate	7,297 (1.7%)
*	Severe	195 (0.0%)
*	Missing	28,125 (6.4%)
Mitral Insufficiency	None	122,180 (27.9%)
*	Trivial	144,803 (33.1%)
*	Mild	117,230 (26.8%)
*	Moderate	30,273 (6.9%)
*	Severe	1,726 (0.4%)
*	Missing	21,232 (4.9%)
Tricuspid Insufficiency	None	128,737 (29.4%)
*	Trivial	173,765 (39.7%)
*	Mild	94,053 (21.5%)
*	Moderate	13,448 (3.1%)
*	Severe	850 (0.2%)
*	Missing	26,591 (6.1%)
Pulmonic Insufficiency	None	248,155 (56.7%)
*	Trivial	98,401 (22.5%)
*	Mild	25,056 (5.7%)
*	Moderate	1,744 (0.4%)
*	Severe	120 (0.0%)
*	Missing	63,968 (14.6%)
Concomitant Tricuspid Repair	No	436,804 (99.9%)
*	Yes	5 (0.0%)
*	Missing	635 (0.1%)

*	Effects	Overall N=437444
Previous CV Interventions	No	288,429 (65.9%)
*	Yes	148,086 (33.9%)
*	Missing	929 (0.2%)
Previous CABG	No	431,229 (98.6%)
*	Yes	5,468 (1.2%)
*	Missing	747 (0.2%)
Previous Valve	No	435,272 (99.5%)
*	Yes	1,365 (0.3%)
*	Missing	807 (0.2%)
Previous PCI	No Prior PCI	304,214 (69.5%)
*	Prior PCI->6 Hours	128,915 (29.5%)
*	Prior PCI-<=6 Hours	3,472 (0.8%)
*	Prior PCI-Timing Missing	250 (0.1%)
*	Missing	593 (0.1%)

The number and descriptive characteristics of patients that are included in the analysis. All eligible operations were included.

* indicates that the cell is left intentionally blank

[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]

We used the same dataset of isolated CABG operations from January 2019 to December 2021 for most of this report. The exceptions are:

1. For comparisons across time periods, and for the empirical validity testing, we calculated hospital-specific mortality estimates using data from January 2016-December 2018 and January 2019-December 2021. Comparisons were limited to participants who participated in STS during both January 2016-December 2018 and January 2019-December 2021.
2. In the individual measure risk prediction model validation section, we cited and reused the sample from the paper published in 2018.

O'Brien SM, Feng L, He X, Xian Y, Jacobs JP, Badhwar V, Kurlansky PA, Furnary AP, Cleveland JC Jr, Lobdell KW, Vassileva C, Wyler von Ballmoos MC, Thourani VH, Rankin JS, Edgerton JR, D'Agostino RS, Desai ND, Edwards FH, Shahian DM. The Society of Thoracic Surgeons 2018 Adult Cardiac Surgery Risk Models: Part 2-Statistical Methods and Results. Ann Thorac Surg. 2018 May;105(5):1419-1428. doi: 10.1016/j.athoracsur.2018.03.003. Epub 2018 Mar 22. PMID: 29577924.

[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]

Whether outcomes measures, and the public reporting and reimbursement programs based on them, should consider socioeconomic (SES) or sociodemographic (SDS) factors (e.g., race, ethnicity, education, income, payer [e.g., Medicare-Medicaid dual eligible status]) is a topic of intense health policy debate [1]. Some argue that in the absence of adjustment for these variables, the outcomes of hospitals that care for a disproportionate percentage of low SES patients will be unfairly disadvantaged, perhaps leading to financial or reputational penalties.

Opponents argue that inclusion of SES factors in risk models may “adjust away” disparities in quality of care, and they advocate the use of stratified analyses instead. They also note that readily available SES factors have often not demonstrated significant impact on outcomes. As part of an NQF pilot project, STS specifically studied dual eligible status in the STS readmission measure [2] and found minimal impact. Finally, even SES proponents agree that these factors make more sense intuitively for some outcomes (e.g., readmission) than others (hospital mortality, complications)—that is, they are context-specific.

In identifying a risk adjustment approach for this measure, and in keeping with the general approach taken for the new STS risk models for the Adult Cardiac Surgery Database [3], we chose to avoid the more philosophical and downstream health policy implications of SES adjustment and based our modeling decisions on empirical findings and consideration of the model's primary intended purpose—to adjust for case mix. Conceptually, our goal was to adjust for all preoperative factors that are independently and significantly associated with outcomes and that vary across STS participants. For example, race will continue to be in our risk models as it has been previously, but not conceptually as a SES indicator. Race has an empirical association with outcomes and has the potential to confound the interpretation of a hospital's outcomes, although we do not know the underlying mechanism (e.g., genetic factors, differential effectiveness of certain medications, rates of certain associated diseases such as diabetes and hypertension).

1. National Academies of Sciences E, and Medicine. Accounting for social risk factors in medicare payment. Washington, DC: The National Academies Press; 2017.
2. Shahian DM, He X, O'Brien SM et al. Development of a clinical registry-based 30-day readmission measure for coronary artery bypass grafting surgery. *Circulation* 2014;130(5):399-409.
3. Shahian DM, Jacobs JP, Badhwar V, Kurlansky PA, Furnary AP, Cleveland JC Jr, Lobdell KW, Vassileva C, Wyler von Ballmoos MC, Thourani VH, Rankin JS, Edgerton JR, D'Agostino RS, Desai ND, Feng L, He X, O'Brien SM. The Society of Thoracic Surgeons 2018 Adult Cardiac Surgery Risk Models: Part 1-Background, Design Considerations, and Model Development. *Ann Thorac Surg.* 2018 May;105(5):1411-1418. doi: 10.1016/j.athoracsur.2018.03.002. Epub 2018 Mar 22. PMID: 29577925.
4. O'Brien SM, Feng L, He X, Xian Y, Jacobs JP, Badhwar V, Kurlansky PA, Furnary AP, Cleveland JC Jr, Lobdell KW, Vassileva C, Wyler von Ballmoos MC, Thourani VH, Rankin JS, Edgerton JR, D'Agostino RS, Desai ND, Edwards FH, Shahian DM. The Society of Thoracic Surgeons 2018 Adult Cardiac Surgery Risk Models: Part 2-Statistical Methods and Results. *Ann Thorac Surg.* 2018 May;105(5):1419-1428. doi: 10.1016/j.athoracsur.2018.03.003. Epub 2018 Mar 22. PMID: 29577924.

[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]

Patient or Encounter-Level (e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements)

Accountable Entity Level (e.g., signal-to-noise analysis)

[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]

See section 2b2 for validity testing of data elements.

For score-level reliability, we re-estimated participant outcomes in a Bayesian statistical framework using Markov Chain Monte Carlo (MCMC) sampling with a non-informative prior distribution. Bayesian analysis allows analysis results to be expressed in terms of probabilities and avoids dependence on large sample approximations. The form of the model is:

$$\text{Prob}(\text{death} | \text{risk score} = x, \text{participant} = j) = \frac{\exp(\alpha + \beta x + \theta_j)}{1 + \exp(\alpha + \beta x + \theta_j)}, \quad \theta_j \sim N(0, \sigma^2) \quad (1)$$

where n is the number of participants, θ_j is the j -th participant's random intercept parameter, and α, β, σ^2 are other unknown parameters to be estimated from the data. The intercept θ_j describes how risk for a patient at participant j compares to a patient with the same STS-predicted risk at other participants. The exponent of the intercept $\exp(\theta_j)$ is the odds ratio comparing participant j to a participant at the median of the performance distribution. The notation $\theta_j \sim N(0, \sigma^2)$ means that the set of participant-specific intercept parameters are assumed to arise from a normal distribution with mean zero and unknown variance, σ^2 . Estimation involves randomly generating M sets of numerical estimates for the set of parameters $\alpha, \beta, \sigma^2, \theta_1, \dots, \theta_n$ and then averaging across the M estimates. For example, an estimate of θ_j is calculated as

$$\hat{\theta}_j = \frac{1}{M} \sum_{i=1}^M \theta_j^{(i)}$$

where $\theta_j^{(i)}$ is the i -th of M randomly generated estimates for θ_j .

Reliability is defined as the proportion of variation in a set of estimates that is due to true between-unit differences (i.e., signal) as opposed to random statistical fluctuations (i.e., noise). A mathematically equivalent definition is the squared correlation between a measurement and the true value. We performed two sets of analyses based on two ways of operationalizing the above squared correlation.

METHOD 1. In the first analysis, we focus attention on the n participants in the STS database and do not attempt to generalize to a hypothetical larger population of participants. Reliability is defined as the square of the

Pearson correlation coefficient (ρ^2) between the set of participant-specific estimates $\hat{\theta}_1, \dots, \hat{\theta}_n$ and the corresponding unknown true values $\theta_1, \dots, \theta_n$, i.e.,

$$\rho^2 = \frac{\left[\sum_{j=1}^n \left(\hat{\theta}_j - \frac{1}{n} \sum_{h=1}^n \hat{\theta}_h \right) \left(\theta_j - \frac{1}{n} \sum_{h=1}^n \theta_h \right) \right]^2}{\sum_{j=1}^n \left(\hat{\theta}_j - \frac{1}{n} \sum_{h=1}^n \hat{\theta}_h \right)^2 \sum_{j=1}^n \left(\theta_j - \frac{1}{n} \sum_{h=1}^n \theta_h \right)^2}.$$

The quantity ρ^2 is estimated as

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

where

$$\rho_{(i)}^2 = \frac{\left[\sum_{j=1}^n \left(\hat{\theta}_j - \frac{1}{n} \sum_{h=1}^n \hat{\theta}_h \right) \left(\theta_j^{(i)} - \frac{1}{n} \sum_{h=1}^n \theta_h^{(i)} \right) \right]^2}{\sum_{j=1}^n \left(\hat{\theta}_j - \frac{1}{n} \sum_{h=1}^n \hat{\theta}_h \right)^2 \sum_{j=1}^n \left(\theta_j^{(i)} - \frac{1}{n} \sum_{h=1}^n \theta_h^{(i)} \right)^2}$$

with $\theta_j^{(i)}$ denoting the value of θ_j on the i -th MCMC iteration and $\hat{\theta}_j = \frac{1}{M} \sum_{i=1}^M \theta_j^{(i)}$ denoting the posterior mean estimate of θ_j . An equal-tailed 95% Bayesian credible interval for ρ^2 was obtained by calculating the empirical 2.5% and 97.5% of $\rho_{(i)}^2$ across the M samples. All calculations were limited to participants with at least 10 eligible operations in the analysis. Thus, n denotes the number of participants with at least 10 eligible operations.

METHOD 2. In the second set of analyses, we assume that STS participants were randomly sampled from a hypothetical infinite population in which outcomes vary according to equation (1) above. Our goal is to estimate reliability across all participants in the population. For a generic participant in the population, let N denote the number of operations it will perform over 3 years, let $X = (x_1, \dots, x_N)$ denote the risk factor profiles of these N cases, let θ denote the participant's intercept that we wish to estimate, let $Y = (Y_1, \dots, Y_N)$ denote the set of mortality indicators for the N patients, and let $\hat{\theta}$ be an estimate of θ . We define reliability as

$$\rho^2 = [\text{corr}(\hat{\theta}, \theta)]^2 = \frac{E[(\hat{\theta} - E\hat{\theta})(\theta - E\theta)]^2}{E[(\hat{\theta} - E\hat{\theta})^2]E[(\theta - E\theta)^2]}$$

where E denotes an average over all possible selections of participants and all possible outcomes of patients at the selected participant. We assume that the distribution of (N, X) in the population is identical to the empirical observed distribution of (N, X) in the current STS database analysis and that $\theta|N, X \sim N(0, \sigma^2)$, where σ^2 is unknown. We assume that $\hat{\theta}$ is calculated using empirical Bayes methodology. Specifically, we assume $\hat{\theta} = \text{argmax}_{\theta} \pi(\theta|N, X, Y, \hat{\alpha}, \hat{\beta}, \hat{\sigma}^2)$ where $\hat{\alpha}, \hat{\beta}, \hat{\sigma}^2$ are the parameter estimates obtained in the current STS database analysis and $\pi(\theta|N, X, Y, \alpha, \beta, \sigma^2)$ is the posterior distribution of θ conditional on the participant's data (N, X, Y) and population parameters $(\alpha, \beta, \sigma^2)$. An implicit assumption is that $\hat{\alpha}, \hat{\beta}, \hat{\sigma}^2$ remain fixed at the values obtained in the current analysis and do not change after observing data from an additional participant. This assumption is made for technical reasons in order to simplify calculations. It is unlikely to have a significant impact because the large number of STS participants makes $\hat{\alpha}, \hat{\beta}, \hat{\sigma}^2$ relatively insensitive to the addition or deletion of

data from any single participant. The quantity ρ^2 defined above depends implicitly on the population parameters α, β, σ^2 which are unobservable. Our calculation allows for the possibility that α, β, σ^2 may differ from the estimates, $\hat{\alpha}, \hat{\beta}, \hat{\sigma}^2$. To make this dependence clear, we write

$$\rho^2 = f(\text{truth} = \alpha, \beta, \sigma^2; \text{estimate} = \hat{\alpha}, \hat{\beta}, \hat{\sigma}^2).$$

A Bayesian estimate of ρ^2 is obtained as

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

where $\rho_{(i)}^2 = f(\text{truth} = \alpha_{(i)}, \beta_{(i)}, \sigma_{(i)}^2; \text{estimate} = \hat{\alpha}, \hat{\beta}, \hat{\sigma}^2)$ with $\alpha_{(i)}, \beta_{(i)}, \sigma_{(i)}^2$ denoting the values of α, β, σ^2 at the i -th MCMC iteration. The function $\rho^2 = f(\text{truth} = \alpha, \beta, \sigma^2; \text{estimate} = \hat{\alpha}, \hat{\beta}, \hat{\sigma}^2)$ does not have a closed form expression. Instead, we approximate it numerically using Monte Carlo sampling with 50,000 iterations within each MCMC iteration.

The above calculation assumes that the distribution of (N, X) in the population is identical to the empirical observed distribution of (N, X) in the current STS database analysis. This yields a global measure of reliability that implicitly averages across the case mix and sample sizes in the STS database. To shed light on how reliability might vary across participants of different case mix and sample sizes, we subsequently performed a set of participant-specific reliability calculations. For each individual STS participant, we asked what reliability would be in a participant population in which all participants have the same sample size and case mix as the STS participant but quality varies according to model (1). We summarized the distribution of participant-specific reliability estimates overall and among participants meeting various sample size thresholds.

Estimation of Misclassification Rates

Additional analyses focused on assessing the measure's classification accuracy. These analyses were based on the following method of assigning participants into performance categories. A participant was labeled as a "high performer" if the 95% probability interval around its intercept fell entirely below the median value of zero. A participant was labeled as a "low performer" if the 95% probability interval around its intercept fell entirely above the median value of zero. The remaining participants were left unclassified as their performance was statistically indistinguishable from the median.

Because a participant's intercept is unobservable, the true misclassification rate of the measure cannot be simply calculated. To infer the likely true misclassification rate, we used MCMC simulations with $M = 4000$ simulation iterations. Let θ_j be the true intercept of the j -th participant and let G_j be an indicator of whether θ_j is above or below the median (1=below median, 0=above median), and let X_j be an indicator of the participant's assigned classification result (1=low performer, 3=high performer, 2=unclassified). Note the X_j is an estimate (observable) whereas θ_j and G_j refer to truth (unobservable). If $X_j = 3$ (high performance) and $G_j = 0$ (low performance) or $X_j = 1$ (low performance) and $G_j = 1$ (high performance) then the participant's true performance is misclassified. To estimate the likely true misclassification rate, we calculated the simulated misclassification rate within each MCMC iteration and then averaged across MCMC iterations. Among participants who were classified as high performers, the true (unobservable) rates of correct classification and misclassification are

$$P_3 = \frac{\sum_{j=1}^n I(X_j = 3) G_j}{\sum_{j=1}^n I(X_j = 3)} \quad \text{and} \quad Q_3 = 1 - P_3 = \frac{\sum_{j=1}^n I(X_j = 3) (1 - G_j)}{\sum_{j=1}^n I(X_j = 3)}$$

where $I(x)$ takes the value 1 if the expression " x " is true and takes the expression " x " is false. Among participants who were classified as low performers, the true (unobservable) rates of correct classification and misclassification are

$$P_1 = \frac{\sum_{j=1}^n I(X_j = 1)(1 - G_j)}{\sum_{j=1}^n I(X_j = 1)} \quad \text{and} \quad Q_1 = 1 - P_1 = \frac{\sum_{j=1}^n I(X_j = 1)G_j}{\sum_{j=1}^n I(X_j = 1)}.$$

Bayesian estimates of P_3 and P_1 (and hence Q_3 and Q_1) were calculated as:

$$\hat{P}_3 = \frac{1}{M} \sum_{i=1}^M P_3^{(i)} \quad \text{and} \quad \hat{P}_1 = \frac{1}{M} \sum_{i=1}^M P_1^{(i)}$$

where

$$P_3^{(i)} = \frac{\sum_{j=1}^n I(X_j = 3)G_j^{(i)}}{\sum_{j=1}^n I(X_j = 3)} \quad \text{and} \quad P_1^{(i)} = \frac{\sum_{j=1}^n I(X_j = 1)(1 - G_j^{(i)})}{\sum_{j=1}^n I(X_j = 1)}$$

with $G_j^{(i)}$ denoting the value of G_j on the i -th of M MCMC iterations. 95% credible intervals for \hat{P}_3 and \hat{P}_1 were obtained by identifying the 2.5% and 97.5% empirical percentiles across the sets of simulated values $P_3^{(1)}, \dots, P_3^{(M)}$ and $P_1^{(1)}, \dots, P_1^{(M)}$.

[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 eCQMs). 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

Quality Improvement (Internal to the specific organization)

[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]

In addition to performance data provided in 1b.2 for Operative Mortality for CABG, we provide the following summary statistics:

1. Aggregate observed rate for the measure among the STS patient-level data by time period;

Risk-Adjusted Event Rate	Observed Rate	Observed Rate
Measure		
Jul15 - Jul16		
Jul16 - Jul17		
0119_CABG_Mortality	2.187%	2.273%

2. Adjusted Odds Ratios for time period 2 (July 2016 – June 2017) vs time period 1 (July 2015 – June 2016): We used a hierarchical logistic regression analysis with a dummy variable for time period and patient factors as fixed effects and participants as random effects. Odds Ratios for time period 2 vs. time period 1 are adjusted by factors included in the STS CABG risk models for each measure.

Adjusted Odds Ratio for
T2: Jul16 - Jun17 vs T1: Jul15 - Jun16
Measure OR Lower 95% CI Upper 95% CI
0119_CABG_Mortality 1.06 1.01 1.12

The slight increase in the event rate from time period 1 (T1) to time period 2 (T2) is reflected in an unfavorable (and statistically significant) adjusted odds ratio for T2 vs. T1.

(Please see Appendix if tables do 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):

0696 The STS CABG 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]

[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 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/John 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

[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]