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 #: 2558

Corresponding Measures:

Measure Title: Hospital 30-day, all-cause, risk-standardized mortality rate (RSMR) following coronary artery bypass graft (CABG) surgery

Measure Steward: Centers for Medicare & Medicaid Services

sp.02. Brief Description of Measure: The measure estimates a hospital-level all-cause, risk standardized mortality rate (RSMR) for patients 65 years and older discharged from the hospital following a qualifying isolated CABG procedure. Mortality is defined as death from any cause within 30 days of the procedure date of an index CABG admission. CMS annually reports the measure for patients who are 65 years or older and enrolled in fee-for-service (FFS) Medicare and hospitalized in non-federal hospitals.

1b.01. Developer Rationale: The goal of this measure is to improve patient outcomes by providing patients, physicians, hospitals, and policy makers with information about hospital-level, risk-standardized mortality rates following hospitalization for a qualifying isolated CABG procedure. Measurement of patient outcomes allows for a broad view of quality of care that encompasses more than what can be captured by individual process-of-care measures. Complex and critical aspects of care, such as communication between providers, prevention of and response to complications, patient safety, and coordinated transitions to the outpatient environment, all contribute to patient outcomes but are difficult to measure by individual process measures. The goal of outcomes measurement is to risk adjust for patients' conditions at the time of hospital admission and then evaluate patient outcomes. This measure was developed to identify institutions whose performance is better or worse than would be expected based on each institution's patient case mix, and therefore promote hospital quality improvement and better inform consumers about care quality.

CABG is a priority area for outcomes measure development because it is a common procedure associated with considerable morbidity, mortality, and health care spending. Between 2013 and 2016, there were 138,785 hospitalizations for CABG surgery among Medicare FFS patients in the U.S [1].

CABG surgeries are costly procedures that account for the majority of major cardiac surgeries performed nationally. In fiscal year 2014, isolated CABG surgeries accounted for almost half (40.59%) of all cardiac surgery hospital admissions in Massachusetts [2]. In 2014, the average Medicare payment was $32,499 for CABG without valve and $45,873 for CABG plus valve surgeries [3].

1. Simoes J, Grady J, DeBuhr J, et al. 2017 Procedure-Specific Measures Updates and Specifications Report Hospital-Level 30-Day Risk-Standardized Mortality Measures.

http://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic/Page/QnetTier3&cid=1163010421830. Accessed March 23, 2018.
2. Massachusetts Data Analysis Center. Adult Coronary Artery Bypass Graft Surgery in the Commonwealth of Massachusetts. https://www.mass.gov/files/documents/2017/12/14/cabg-fy2014.pdf. Accessed March 23, 2018.
3. Pennsylvania Health Care Cost Containment Council. Hospital Medicare Payment. http://www.phc4.org/reports/cabg/16/docs/Hospital%20Medicare%20Payment.pdf. Accessed March 23, 2018.

---

sp.12. Numerator Statement: The outcome for this measure is 30-day all-cause mortality. Mortality is defined as death for any reason within 30 days of the procedure date from the index admission for patients 65 and older discharged from the hospital after undergoing isolated CABG surgery.

sp.14. Denominator Statement:

This claims-based measure is used for a cohort of patients aged 65 years or older.

The cohort includes admissions for patients who receive a qualifying isolated CABG procedure (see the attached Data Dictionary) and with a complete claims history for the 12 months prior to admission.

For patients with more than one qualifying CABG surgery admission in the measurement period, the first CABG admission is selected for inclusion in the measure and the subsequent CABG admission(s) are excluded from the cohort.

sp.16. Denominator Exclusions:

The CABG surgery mortality measure excludes index admissions for patients:

1. With inconsistent or unknown vital status or other unreliable demographics (age and gender) data; or,
2. Discharged against medical advice (AMA).

For patients with more than one qualifying CABG surgery admission in the measurement period, the first CABG admission is selected for inclusion in the measure and the subsequent CABG admission(s) are excluded from the cohort.

Measure Type: Outcome

sp.28. Data Source:

Other (specify)

Enrollment data

Claims

sp.07. Level of Analysis:

Facility

IF Endorsement Maintenance – Original Endorsement Date: 2014-11-12 05:24 PM

Most Recent Endorsement Date: 10/26/2018 12:01:49 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]

Yes

[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]

The goal of this measure is to improve patient outcomes by providing patients, physicians, hospitals, and policy makers with information about hospital-level, risk-standardized mortality rates following hospitalization for a qualifying isolated CABG procedure. Measurement of patient outcomes allows for a broad view of quality of care that encompasses more than what can be captured by individual process-of-care measures. Complex and critical aspects of care, such as communication between providers, prevention of and response to complications, patient safety, and coordinated transitions to the outpatient environment, all contribute to patient outcomes but are difficult to measure by individual process measures. The goal of outcomes measurement is to risk adjust for patients' conditions at the time of hospital admission and then evaluate patient outcomes. This measure was developed to identify institutions whose performance is better or worse than would be expected based on each institution's patient case mix, and therefore promote hospital quality improvement and better inform consumers about care quality.

CABG is a priority area for outcomes measure development because it is a common procedure associated with considerable morbidity, mortality, and health care spending. Between 2013 and 2016, there were 138,785 hospitalizations for CABG surgery among Medicare FFS patients in the U.S [1].

CABG surgeries are costly procedures that account for the majority of major cardiac surgeries performed nationally. In fiscal year 2014, isolated CABG surgeries accounted for almost half (40.59%) of all cardiac surgery hospital admissions in Massachusetts [2]. In 2014, the average Medicare payment was $32,499 for CABG without valve and $45,873 for CABG plus valve surgeries [3].

1. Simoes J, Grady J, DeBuhr J, et al. 2017 Procedure-Specific Measures Updates and Specifications Report Hospital-Level 30-Day Risk-Standardized Mortality Measures. http://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic/Page/QnetTier3&cid=1163010421830. Accessed March 23, 2018.
2. Massachusetts Data Analysis Center. Adult Coronary Artery Bypass Graft Surgery in the Commonwealth of Massachusetts. https://www.mass.gov/files/documents/2017/12/14/cabg-fy2014.pdf. Accessed March 23, 2018.
3. Pennsylvania Health Care Cost Containment Council. Hospital Medicare Payment. http://www.phc4.org/reports/cabg/16/docs/Hospital%20Medicare%20Payment.pdf. Accessed March 23, 2018.

[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]

Variation in mortality rates indicates opportunity for improvement. We conducted analyses using data from July 1, 2013 to June 30, 2016 Medicare claims data (n= 138,66 admissions from 1,185 hospitals) and reported hospital-level RSRMs having a mean of 3.3% (SD=0.9) and a range of 1.3% - 7.4%. The median RSRR is 3.1% (20th and 70th percentiles are 2.6% and 3.6%, respectively). The distribution of RSRMs across hospitals is shown below:

Distribution of Hospital CABG RSMRs over Different Time Periods
Results for each data year

Characteristic//07/2013-06/2014//07/2014-06/2015//07/2015-06/2016//07-2013-06/2016
Number of Hospitals// 1,158 // 1,150 // 1,151 // 1,185
Number of Admissions// 46,279 // 46,123 // 46,259 // 138,661
Mean (SD)// 3.2 (0.5) // 3.4 (0.7) // 3.2 (0.7) // 3.3 (0.9)
Range (min. – max.)// 1.9 – 6.0 // 1.8 – 6.7 // 1.4 – 6.8 // 1.3 – 7.4
Minimum// 1.9 // 1.8 // 1.4 // 1.3
10th percentile// 2.7 // 2.6 // 2.5 // 2.3
20th percentile// 2.8 // 2.9 // 2.7 // 2.6
30th percentile// 2.9 // 3.0 // 2.8 // 2.8
40th percentile// 3.0 // 3.1 // 2.9 // 3.0
50th percentile// 3.1 // 3.2 // 3.0 // 3.1
60th percentile// 3.2 // 3.3 // 3.1 // 3.3
70th percentile// 3.3 // 3.6 // 3.4 // 3.6
80th percentile// 3.5 // 3.8 // 3.7 // 3.9
90th percentile// 3.8 // 4.2 // 4.1 // 4.4
Maximum// 6.0 // 6.7 // 6.8 // 7.4

[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]

Distribution of CABG RSMRs by Proportion of Dual Eligible Patients:
Data Source: Medicare FFS claims
Dates of Data: July 2013 through June 2016

Characteristic// Hospitals with a low proportion (=5.6%) Dual Eligible patients//Hospitals with a high proportion (=13.4%) Dual Eligible patients
Number of Measured Entities (Hospitals)// 260 // 260
Number of Patients// 123,442 patients in low-proportion hospitals// 13,628 in high-proportion hospitals
Maximum// 7.4 // 7.2
90th percentile// 4.2// 4.8
75th percentile// 3.5 // 3.9
Median (50th percentile)// 3.1 // 3.2
25th percentile// 2.6 // 2.7
10th percentile// 2.3 // 2.4
Minimum // 1.3 // 1.5

Distribution of CABG RSMRs by Proportion of African-American Patients:
Data Source: Medicare FFS claims
Dates of Data: July 2013 through June 2016

Characteristic// Hospitals with a low Proportion (=0.7%) African-American patients//Hospitals with a high proportion (=7.1%) African-American patients
Number of Measured Entities (Hospitals)// 259 // 266
Number of Patients//131,354 patients in low-proportion hospitals// 7,307 in high-proportion hospitals
Maximum// 6.0 // 6.4
90th percentile// 4.5 // 4.6
75th percentile// 3.8 // 3.8
Median (50%)// 3.2 // 3.1
25th percentile// 2.7 // 2.7
10th percentile// 2.4 // 2.3
Minimum// 1.6 // 1.5

Distribution of CABG RSMRs by Proportion of Patients with AHRQ SES Index Scores Equal to or Below 42.6:
Data Source: Medicare FFS claims and The American Community Survey (2008-2012) data
Dates of Data: July 2013 through June 2016

Characteristic//Hospitals with low proportion of patients with AHRQ SES index score equal to or below 42.6 (=8.8%)//Hospitals with high proportion of patients with AHRQ SES index score equal to or below 42.6 (=26.8%)
Number of Measured Entities (Hospitals)// 259 // 259
Number of Patients// 112,666 patients in hospitals with low proportion of patients with AHRQ SES index score equal to or below 42.6 // 25,995 patients in hospitals with high proportion of patients with AHRQ SES index score equal to or below 42.6
Maximum// 7.4 // 7.2
90th percentile// 4.1 // 4.8
75th percentile// 3.4 // 4.1
Median (50th percentile)// 2.9 // 3.5
25th percentile// 2.5 // 2.8
10th percentile// 2.1// 2.4
Minimum // 1.3 // 1.7

[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]

Please see section spma.02 below for detailed updates to the measure specifications since the last submission.

[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]

Overall, measure updates were limited to coding updates. There were no substantial updates to this measure since the last submission.

[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]

Hospital 30-day, all-cause, risk-standardized mortality rate (RSMR) following coronary artery bypass graft (CABG) surgery

[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]

The measure estimates a hospital-level all-cause, risk standardized mortality rate (RSMR) for patients 65 years and older discharged from the hospital following a qualifying isolated CABG procedure. Mortality is defined as death

from any cause within 30 days of the procedure date of an index CABG admission. CMS annually reports the measure for patients who are 65 years or older and enrolled in fee-for-service (FFS) Medicare and hospitalized in non-federal hospitals.

[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: Coronary Artery Disease

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

[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]

Elderly (Age >= 65)

[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]

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://qualitynet.org/inpatient/measures/mortality/methodology

[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]

Available in attached Excel or csv file

[Response Ends]

Attachment: 2558_2020 CABG Readmissions_final.xlsx

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]

The outcome for this measure is 30-day all-cause mortality. Mortality is defined as death for any reason within 30 days of the procedure date from the index admission for patients 65 and older discharged from the hospital after undergoing isolated CABG surgery.

[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]

In the current publicly reported measure, we identify deaths for Medicare FFS patients 65 years or older in the Medicare Enrollment Database (EDB). We use the date of death in the enrollment database (EDB) which is derived from the Social Security Administration and has been verified.

Outcome Attribution:

Attribution of the outcome in situations where a patient has multiple contiguous admissions, at least one of which involves a qualifying isolated CABG procedure is as follows:

1. If a patient undergoes a CABG procedure in the first hospital and is then transferred to a second hospital where there is no CABG procedure, the mortality outcome is attributed to the first hospital performing the index CABG procedure and the 30-day window starts with the date of index CABG procedure.

Rationale: A transfer following CABG is most likely due to a complication of the index procedure and that care provided by the hospital performing the CABG procedure likely dominates mortality risk even among transferred patients.

1. If a patient is admitted to a first hospital but does not receive a CABG procedure there and is then transferred to a second hospital where a CABG is performed, the mortality outcome is attributed to the second hospital performing the index CABG procedure and the 30-day window starts with the date of index CABG procedure.

Rationale: Care provided by the hospital performing the CABG procedure likely dominates mortality risk.

1. If a patient undergoes a CABG procedure in the first hospital and is transferred to a second hospital where another CABG procedure is performed, the mortality outcome is attributed to the first hospital performing the index (first) CABG procedure and the 30-day window starts with the date of index CABG procedure.

Rationale: A transfer following CABG is most likely due to a complication of the index procedure, and care provided by the hospital performing the index CABG procedure likely dominates mortality risk even among transferred patients.

[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]

This claims-based measure is used for a cohort of patients aged 65 years or older.

The cohort includes admissions for patients who receive a qualifying isolated CABG procedure (see the attached Data Dictionary) and with a complete claims history for the 12 months prior to admission.

For patients with more than one qualifying CABG surgery admission in the measurement period, the first CABG admission is selected for inclusion in the measure and the subsequent CABG admission(s) are excluded from the cohort.

[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]

The measure includes index admissions for patients:

1. Having a qualifying isolated CABG surgery during the index admission;
2. Enrolled in Medicare fee-for-service (FFS) Part A and Part B for the 12 months prior to the date of the index admission, and enrolled in Part A during the index admission; and,
3. Aged 65 or over.

Isolated CABG surgeries are defined as those CABG procedures performed without the following concomitant valve or other major cardiac, vascular, or thoracic procedures:

- Valve procedures;
- Atrial and/or ventricular septal defects;
- Congenital anomalies;
- Other open cardiac procedures;
- Heart transplants;
- Aorta or other non-cardiac arterial bypass procedures;
- Head, neck, intracranial vascular procedures; or,
- Other chest and thoracic procedures

International Classification of Disease, 10th Revision (ICD-10) codes used to define the cohort are listed in the attached Data Dictionary.

[Response Ends]

sp.17. Describe the denominator exclusions.

Brief narrative description of exclusions from the target population.

[Response Begins]

The CABG surgery mortality measure excludes index admissions for patients:

1. With inconsistent or unknown vital status or other unreliable demographics (age and gender) data; or,
2. Discharged against medical advice (AMA).

For patients with more than one qualifying CABG surgery admission in the measurement period, the first CABG admission is selected for inclusion in the measure and the subsequent CABG admission(s) are excluded from the cohort.

[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]

The CABG surgery mortality measure excludes index admissions for patients:

1. With inconsistent or unknown vital status or other unreliable demographics (age and gender) data.

   Rationale: We do not include stays for patients where the age (indicated in the claim) is greater than 115, where the gender (indicated in the claim) is not coded as male or female, where the admission date (indicated in the claim) is after the date of death in the Medicare Enrollment Database, or where the date of death (in the Medicare Enrollment Database) occurs before the date of discharge but the patient was discharged alive (indicated in the claim).

2. Discharged against medical advice (AMA).

   Rationale: Providers did not have the opportunity to deliver full care and prepare the patient for discharge. This information is taken from the discharge disposition in the claim.

3. With more than one qualifying CABG surgery admission in the measurement period.

   Rationale: CABG procedures are expected to last for several years without the need for revision or repeat revascularization. A repeat CABG procedure during the measurement period likely represents a complication of the original CABG procedure and is a clinically more complex and higher risk surgery. Therefore, we select the first CABG surgery admission for inclusion in the measure and exclude subsequent CABG surgery admissions from the cohort.

[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]

Not applicable; this measure is not stratified.

[Response Ends]

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

[Response Begins]

No

[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]

The measure estimates hospital-level 30-day all-cause RSMRs for CABG surgery using a hierarchical logistic regression model. In brief, the approach simultaneously models data at the patient and hospital levels to account for variance in patient outcomes within and between hospitals (Normand and Shahian, 2007). At the patient level, it models the log-odds of mortality within 30 days of the procedure date using age, sex, selected clinical covariates, and a hospital-specific effect. At the hospital level, the approach models the hospital-specific effects as arising from a normal distribution. The hospital effect represents the underlying risk of mortality at the hospital, after

accounting for patient risk. The hospital-specific effects are given a distribution to account for the clustering (nonindependence) of patients within the same hospital (Normand and Shahian, 2007). If there were no differences among hospitals, then after adjusting for patient risk, the hospital effects should be identical across all hospitals.

The RSMR is calculated as the ratio of the number of "predicted" deaths to the number of "expected" deaths at a given hospital, multiplied by the national observed mortality rate. For each hospital, the numerator of the ratio is the number of deaths within 30 days predicted based on the hospital's performance with its observed case mix, and the denominator is the number of deaths expected based on the nation's performance with that hospital's case mix. This approach is analogous to a ratio of "observed" to "expected" used in other types of statistical analyses. It conceptually allows a particular hospital's performance, given its case mix, to be compared to an average hospital's performance with the same case mix. Thus, a lower ratio indicates lower-than-expected mortality rates or better quality, while a higher ratio indicates higher-than-expected mortality rates or worse quality.

The "predicted" number of deaths (the numerator) is calculated by using the coefficients estimated by regressing the risk factors and the hospital-specific effect on the risk of mortality. The estimated hospital-specific effect is added to the sum of the estimated regression coefficients multiplied by the patient characteristics. The results are log transformed and summed over all patients attributed to a hospital to get a predicted value. The "expected" number of deaths (the denominator) is obtained in the same manner, but a common effect using all hospitals in our sample is added in place of the hospital-specific effect. The results are log transformed and summed over all patients in the hospital to get an expected value. To assess hospital performance for each reporting period, we re-estimate the model coefficients using the years of data in that period.

This calculation transforms the ratio of predicted over expected into a rate that is compared to the national observed mortality rate.

The hierarchical logistic regression models are described fully in the original methodology report (Suter et al. 2012).

References:

Normand S-LT, Shahian DM. 2007. Statistical and Clinical Aspects of Hospital Outcomes Profiling. Stat Sci 22(2): 206-226.

Suter L, Wang C, Araas M, et al. Hospital-Level 30-day All-Cause Mortality Following Coronary Artery Bypass Graft Surgery; Updated Measure Methodology Report. 2012.

[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. This measure is not based on a sample or survey.

[Response Ends]

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

[Response Begins]

Claims

Other (specify)

[Other (specify) Please Explain]

Enrollment 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]

Data sources for the Medicare FFS measure:

Medicare Part A inpatient and Part B outpatient claims: This data source contains claims data for FFS inpatient and outpatient services including: Medicare inpatient hospital care, outpatient hospital services, as well as inpatient and outpatient physician claims for the 12 months prior to an index admission.

Medicare Enrollment Database (EDB): This database contains Medicare beneficiary demographic, benefit/coverage, and vital status information. This data source was used to obtain information on several inclusion/exclusion indicators such as Medicare status on admission as well as vital status. These data have previously been shown to accurately reflect patient vital status (Fleming et al., 1992).

Reference:

Fleming C., Fisher ES, Chang CH, Bubolz D, Malenda J. Studying outcomes and hospital utilization in the elderly: The advantages of a merged data base for Medicare and Veterans Affairs Hospitals. Medical Care. 1992; 30(5): 377-91.

[Response Ends]

sp.32. Provide the data collection instrument.

[Response Begins]

No data collection instrument provided

[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]

Yes - Additional risk adjustment analysis is included

[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]

Claims

Other (specify)

[Other (specify) Please Explain]

Medicare Enrollment Data (including the Master Beneficiary Summary File); American Community Survey (ACS)

[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]

The datasets used for testing included Medicare Parts A and B claims as well as the Medicare Enrollment Database (EDB). To assess socioeconomic factors, we used census as well as claims data (dual eligible status obtained through enrollment data; Agency for Healthcare Research and Quality (AHRQ) socioeconomic status (SES) index score calculated from the American Community Survey). Race variable (Black) was obtained through the Medicare Enrollment Database. The dataset used varies by testing type; see Section 2a.07 for details.

[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]

The dates used vary by testing type; see Section 2a.07 for details.

[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]

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]

For this measure, hospitals are the measured entities. All non-federal, acute inpatient US hospitals (including territories) with Medicare Fee-for-Service (FFS) beneficiaries aged 65 years or over are included. The number of measured entities (hospitals) varies by testing type; see Section 2a.07 for details.

[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]

The number of admissions/patients varies by testing type; see Section 2a.07 for details.

[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]

The datasets, dates, number of measured hospitals, and number of admissions used in each type of testing are shown in Table 1.

Dataset	Applicable Section	Description of Dataset
Development and Validation Datasets (Medicare Fee-For-Service Administrative Claims Data) (Referred to as "Dataset 2" in the previous submission)	2b.20 Statistical Risk Model Discrimination Statistics 2b.28 Statistical Risk Model Calibration Statistics 2b.30 Risk Adjustment/Stratification	Dates of Data: January 1, 2008 – December 31, 2010 Medicare Part A Inpatient and Outpatient and Part B Outpatient claims Number of Admissions: 173,291 Patient Descriptive Characteristics: average age=81.0, % male=46.1 Number of Measured Hospitals: 1,170

Dataset	Applicable Section	Description of Dataset
Previous Submission: 2017 Public Reporting Dataset (referred to in the previous submission as "Dataset 1")	2b.01 Validity Testing 2b.05 Meaningful Differences 2a.09 Reliability Testing 2b.16 Testing of Measure Exclusions 2b.20 Statistical Risk Model Discrimination Statistics 2b.30 Risk Adjustment/Stratification	Dates of Data: July 1, 2013 – June 30, 2016 (2017 public reporting cohort) Number of Admissions: 138,661 Patient Descriptive Characteristics: average age=73.7, % male=71.7 Number of Measured Hospitals: 1,185
Current Submission: Fall 2020 Endorsement Maintenance Testing Dataset (Medicare Fee-For-Service Administrative Claims Data (July 1, 2016 – June 30, 2019)	2b.05 Meaningful Differences 2a.09 Reliability Testing 2b.011 Validity Testing 2b.16 Testing of Measure Exclusions 2b.20 Statistical Risk Model Discrimination Statistics 2b.30 Risk Adjustment/Stratification	Dates of Data: July 1, 2016 – June 30, 2019 Number of Admissions: 135,292 Patient Descriptive Characteristics: average age=73.5, % male=73.2 Number of Measured Hospitals: 1,163
The American Community Survey (ACS)	Section 2b.30: Risk adjustment/Stratification	Original development: 2008-2012 Current submission: Dates of Data: 2013-2017 We used the AHRQ SES index score derived from the American Community Survey (2013-2017) to study the association between the 30-day mortality outcome and social risk factors. The AHRQ SES index score is based on beneficiary 9-digit zip code level of residence and incorporates 7 census variables found in the American Community Survey.
Master Beneficiary Summary File (MBSF)	Section 2b.30: Risk adjustment/Stratification	Current submission: Dates of Data: July 2016 – June 2019 We used dual eligible status (for Medicare and Medicaid) derived from the MBSF to study the association between the 30-day measure outcome and dual-eligible status.

Table 1: Dataset descriptions

[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]

We selected SES variables to analyze after reviewing the literature, developing our conceptual model, and examining available national data sources. The causal pathways for SES variable selection are described below in Section 2b.23.The SES variables used for analysis were:

- **Dual eligible status**: Dual eligible status (i.e., enrolled in both Medicare and Medicaid) patient-level data is obtained from the CMS Master Beneficiary Summary File (MBSF).

  Following guidance from ASPE (ASPE 2016; ASPE 2020), NQF (NQF, 2021), and a body of literature demonstrating differential health care and health outcomes among dual eligible patients, we identified dual eligibility as a key variable. We recognize that Medicare-Medicaid dual eligibility has limitations as a proxy for patients' income or assets because it does not provide a range of results and is only a dichotomous outcome. However, the threshold for over 65-year-old Medicare patients is valuable, as it takes into account both income and assets and is consistently applied across states for the older population. We acknowledge that it is important to test a wider variety of social risk factors including key variables such as education and poverty level; therefore, we also tested a validated composite based on census data linked to as small a geographic unit as possible.

- **AHRQ-validated SES index score** (summarizing the information from the following seven variables): percentage of people in the labor force who are unemployed, percentage of people living below poverty level, median household income, median value of owner-occupied dwellings, percentage of people ≥25 years of age with less than a 12th grade education, percentage of people ≥25 years of age completing ≥4 years of college, and percentage of households that average ≥1 people per room.

Finally, we selected the AHRQ SES index score because it is a well-validated variable that describes the average SES of people living in defined geographic areas (Bonito et al., 2008). Its value as a proxy for patient-level information is dependent on having the most granular-level data with respect to communities that patients live in. We considered the area deprivation index (ADI) among many other potential indicators when we initially evaluated the impact of SDS indicators. We ultimately did not include the ADI at the time, partly due to the fact that the coefficients used to derive ADI had not been updated for many years. Recently, the coefficients for ADI have been updated and therefore we compared the ADI with the AHRQ SES Index and found them to be highly correlated. In this submission, we present analyses using the census block level, the most granular level possible using American Community Survey (ACS) data. A census block group is a geographical unit used by the US Census Bureau which is between the census tract and the census block. It is the smallest geographical unit for which the bureau publishes sample data. The target size for block groups is 1,500 and they typically have a population of 600 to 3,000 people. We used 2013-2017 ACS data and mapped patients' 9-digit ZIP codes via vendor software to the census block group level. Given the variation in cost of living across the country, the median income and median property value components of the AHRQ SES Index were adjusted by regional price parity values published by the Bureau of Economic Analysis (BEA). This provides a better marker of low SES neighborhoods in high expense geographic areas. We then calculated an AHRQ SES Index score for census block groups that can be linked to 9-digit ZIP codes. We used the percentage of patients with an AHRQ SES index score equal to or below 46 to define the lowest quartile of the AHRQ SES Index.

- **Race (Black) (added for current submission)**

Data source: Medicare enrollment database

We used the Medicare enrollment database to identify the patient-level race variable (Black) that we used in these analyses. The Black variable has been shown to be reliable for use in this dataset (Waldo, 2004)

References:

Boan AD, Feng WW, Ovbiagele B, et al. Persistent racial disparity in stroke hospitalization and economic impact in young adults in the buckle of stroke belt. Stroke; a journal of cerebral circulation. Jul 2014;45(7):1932-1938.

Bonito A, Bann C, Eicheldinger C, Carpenter L. Creation of new race-ethnicity codes and socioeconomic status (SES) indicators for Medicare beneficiaries. Final Report, Sub-Task. 2008;2.

Clark CJ, Guo H, Lunos S, et al. Neighborhood cohesion is associated with reduced risk of stroke mortality. Stroke; a journal of cerebral circulation. May 2011;42(5):1212-1217.

Department of Health and Human Services, Office of the Assistant Secretary of Planning and Evaluation (ASPE). Report to Congress: Social Risk factors and Performance Under Medicare's Value-based Payment Programs. 2016; https://aspe.hhs.gov/pdf-report/report-congress-social-risk-factors-and-performance-under-medicares-value-based-purchasing-programs. Accessed November 10, 2019.

Department of Health and Human Services, Office of the Assistant Secretary of Planning and Evaluation (ASPE). Second Report to Congress: Social Risk Factors and Performance in Medicare's Value-based Purchasing Programs. 2020; https://aspe.hhs.gov/reports/second-report-congress-social-risk-medicares-value-based-purchasing-programs Accessed July 2, 2020.

Glymour MM, Kosheleva A, Boden-Albala B. Birth and adult residence in the Stroke Belt independently predict stroke mortality. Neurology. Dec 1 2009;73(22):1858-1865.

Howard VJ, Kleindorfer DO, Judd SE, et al. Disparities in stroke incidence contributing to disparities in stroke mortality. Ann Neurol 2011;69:619–627.

Khan JA, Casper M, Asimos AW, et al. Geographic and sociodemographic disparities in drive times to Joint Commission-certified primary stroke centers in North Carolina, South Carolina, and Georgia. Preventing chronic disease. Jul 2011;8(4):A79.

National Quality Forum. Driving Measurable Health Improvements Together Developing and Testing Risk Adjustment Models for Social and Functional Status-Related Risk within Healthcare Performance Measurement August 2021. Accessed at: https://www.qualityforum.org/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=96087

Pedigo A, Seaver W, Odoi A. Identifying unique neighborhood characteristics to guide health planning for stroke and heart attack: fuzzy cluster and discriminant analyses approaches. PloS one. 2011;6(7):e22693.

van Oeffelen AA, Agyemang C, Bots ML, et al. The relation between socioeconomic status and short-term mortality after acute myocardial infarction persists in the elderly: results from a nationwide study. European journal of epidemiology. Aug 2012; 27(8):605-613.

Waldo DR. Accuracy and Bias of Race/Ethnicity Codes in the Medicare Enrollment Database. Health Care Financing Review. 2004;26(2). https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4194866/

[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]

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]

Previous Submission

Split-sample reliability

The reliability of a measurement is the degree to which repeated measurements of the same entity agree with each other. For measures of hospital performance, the measured entity is naturally the hospital, and reliability is the extent to which repeated measurements of the same hospital give similar results. In line with this thinking, our approach to assessing reliability is to consider the extent to which assessments of a hospital using different but randomly selected subsets of patients produces similar measures of hospital performance. That is, we take a "test-retest" approach in which hospital performance is measured once using a random subset of patients, then measured again using a second random subset exclusive of the first, and finally compare the agreement between the two resulting performance measures across hospitals (Rousson et al., 2002).

For test-retest reliability, we combined index admissions from successive measurement periods into one dataset, randomly sampled half of the patients within each hospital, calculated the measure for each hospital, and repeated the calculation using the second half. Thus, each hospital is measured twice, but each measurement is made using an entirely distinct set of patients. To the extent that the calculated measures of these two subsets agree, we have evidence that the measure is assessing an attribute of the hospital, not of the patients. As a metric of agreement, we calculated the intra-class correlation coefficient (ICC) (Shrout and Fleiss, 1979). Specifically, we used Dataset 1 split sample and calculated the RSMR for each hospital for each sample. The agreement of the two RSMRs was quantified for hospitals using the intra-class correlation as defined by ICC (2, 1) by Shrout and Fleiss (1979).

Using two independent samples provides a stringent estimate of the measure's reliability, in comparison to using two random, but potentially overlapping, samples which would exaggerate the agreement. Moreover, because our final measure is derived using hierarchical logistic regression, and a known property of hierarchical logistic regression models is that smaller-volume hospitals contribute less ´signal´, a split sample using a single measurement period would introduce extra noise. This leads to an underestimate in the actual test-retest reliability that would be achieved if the measure were reported using the full measurement period, as evidenced by the Spearman Brown prophecy formula (Spearman, 1910; Brown, 1910). We use this to estimate the reliability of the measure if the whole cohort were used, based on an estimate from half the cohort.

Test-retest reliability is considered the lower bound of any reliability estimate (Yu, Mehrotra, and Adams, 2013). While it is the most relevant metric from the perspective of measure reliability, it is also meaningful to consider the separate notion of "unit" reliability, that is, the reliability with which individual units (here, hospitals) are measured. Therefore, we also use the approach used by Adams and colleagues to calculate reliability for this measure (2010). Because this metric has been reported for other measures in other contexts (see e.g., Adams et al 2010), and to provide an additional, complementary metric, we also report this average unit reliability.

Current Submission

Signal-to-Noise Reliability

For this updated submission we estimated the signal to noise reliability (facility-level reliability), which is the reliability with which individual units (hospitals) are measured.

We used the formula presented by Adams and colleagues (2010) to calculate facility-level reliability. Where facility-to-facility variance is estimated from the hierarchical logistic regression model, n is equal to each facility's observed case size, and the facility error variance is estimated using the variance of the logistic distribution ($\pi^2/3$). The facility-level reliability testing is limited to facilities with at least 25 admissions for public reporting.

Signal to noise reliability scores can range from 0 to 1. A reliability of zero implies that all the variability in a measure is attributable to measurement error. A reliability of one implies that all the variability is attributable to real difference in performance.

References:

Adams J, Mehrota, A, Thoman J, McGlynn, E. (2010). Physician cost profiling – reliability and risk of misclassification. NEJM, 362(11): 1014-1021.

Brown, W. (1910). Some experimental results in the correlation of mental abilities. British Journal of Psychology, 3, 296–322.

Rousson V, Gasser T, Seifert B. Assessing intrarater, interrater and test–retest reliability of continuous measurements. Statistics in Medicine 2002;21:3431-3446.

Shrout P, Fleiss J. Intraclass correlations: uses in assessing rater reliability. Psychological Bulletin 1979;86:420-428.

Spearman, Charles, C. (1910). Correlation calculated from faulty data. British Journal of Psychology, 3, 271-295

Yu, H, Mehrota, A, Adams J. (2013). Reliability of utilization measures for primary care physician profiling. Healthcare, 1, 22-29.

[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]

Previous submission

Split-Sample Reliability

There were 138,661 admissions in the 2017 public reported CABG mortality measure (**Dataset 1**), with 69,040 in one sample and 69,621 in the other randomly selected sample. The agreement between the two RSMRs for each hospital was 0.35.

Signal-to-Noise Reliability

Please note that the above split-sample reliability represents the lower bound of any reliability estimate of this measure. Using the approach by Adams et al (2010), we found the mean reliability score to be 0.851. This is considered to be high (Yu, Mehrotra, and Adams, 2013).

Current Submission

Below we provide volume-stratified distribution of signal-to-noise reliability results using the Fall 2022 Endorsement Maintenance Dataset

Number of Admissions	# hospitals	Mean	Std	Min	5th Pctl	10th Pctl	25th Pctl	Median	75th Pctl	90th Pctl	95th Pctl	Max
>=1	1,163	0.765	0.205	0.055	0.291	0.499	0.711	0.836	0.901	0.938	0.953	0.98

Number of Admissions	# hospitals	Mean	Std	Min	5th Pctl	10th Pctl	25th Pctl	Median	75th Pctl	90th Pctl	95th Pctl	Max
>=25	1,002	0.834	0.095	0.594	0.645	0.69	0.769	0.86	0.909	0.941	0.954	0.98

Table 2: Distribution of signal-to-noise reliability

References:

Adams J, Mehrota, A, Thoman J, McGlynn, E. (2010). Physician cost profiling – reliability and risk of misclassification. NEJM, 362(11): 1014-1021.

Simoes J, Grady J, DeBuhr J et al., 2017 Procedure-Specific Measure Updates and Specifications Report Hospital-Level 30-Day Risk-Standardized Mortality Measure Isolated Coronary Artery Bypass Graft (CABG) Surgery – Version 4.0. Available at: https://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier2&cid=1163010398556

Yu, H, Mehrota, A, Adams J. (2013). Reliability of utilization measures for primary care physician profiling. Healthcare, 1, 22-29.

[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]

Previous Submission

The ICC demonstrates fair agreement in measure score reliability.

The ICC[2,1] is a conservative measure of test-retest reliability because it assumes that the multiple measurements are drawn from a larger sample of tests, and that the measured providers are drawn from a larger sample of providers. Given the conservative nature of the ICC[2,1] and the complex constructs of risk-adjusted outcome measures, a lower reliability score is expected.

Our test-retest reliability score of 0.35 represents the lower bound of any reliability estimate. Using the approach used by Adams et al (2010), we obtained mean reliability score of 0.851. This pattern was also observed by Yu, Mehrotra and Adams (2013). For example, they found mean reliability for a PCP visits utilization measure to be 0.94 using the approach used by Adams and colleagues (2010), although the rest-retest reliability score was 0.68. Taking together these results indicate that there is sufficient reliability in the measure score.

Current Submission

The median signal-to-noise reliability for hospitals with at least 25 cases (0.86) is high, and sufficient for a publicly reported quality measure.

[Response Ends]

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

[Response Begins]

Accountable Entity Level (e.g. hospitals, clinicians)

Empirical validity testing

Systematic assessment of face validity of performance measure score as an indicator of quality or resource use (i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance)

[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]

Previous submission

Measure validity is demonstrated through prior validity testing done on our other claims-based measures, through use of established measure development guidelines, and by systematic assessment of measure face validity by a TEP of national experts and stakeholder organizations.

Face Validity as Determined by TEP

To systematically assess face validity, we surveyed the TEP and asked each member to rate the following statement using a six-point scale (1=Strongly Disagree, 2=Moderately Disagree, 3=Somewhat Disagree, 4=Somewhat Agree, 5= Moderately Agree, and 6=Strongly Agree): "The mortality rates obtained from the mortality measure as specified will provide an accurate reflection of quality."

Measure Score Validity -Validity Indicated by Established Measure Development Guidelines

We developed this measure in consultation with national guidelines for publicly reported outcome measures, with outside experts, and with the public. The measure is consistent with the technical approach to outcome measurement set forth in NQF guidance for outcome measures, CMS MMS guidance, and the guidance articulated in the American Heart Association scientific statement, "Standards for Statistical Models Used for Public Reporting of Health Outcomes" (Krumholz, Brindis, et al. 2006; NQF 2010).

Validation of the Administrative Risk-Adjustment Model using clinical registry data

To validate the administrative risk-adjustment model, we calculated hospital-level, RSMRs using the claims-based CABG mortality measure risk model and a risk model created using clinical registry data in a common cohort of isolated CABG patients (2008-2010) from the New York State Cardiac Surgery Reporting System (CSRS) from the New York Department of Health and compared the results. We matched claims from the 2008-2010 data sets to the 2008-2011 NY Registry data.

We measured the correlation between the two sets of results at the hospital level. In addition, we used a bootstrapping approach similar to that used for public reporting of the AMI, heart failure and pneumonia mortality measures to categorize hospital performance as better, worse or no different than the average hospital observed mortality rate. The bootstrapping algorithm used is described in the methodology report.

We then performed a reclassification analysis to determine how many hospitals might be reclassified to a different performance category if assessed by the administrative model as compared to the registry model. In order to isolate differences due to the method of risk adjustment, both measures were calculated in the same cohort of patients, used the same outcome definition (30-day all-cause mortality defined by administrative claims data) and a consistent approach to risk-adjustment modeling (the hierarchical logistic regression model approach used in CMS's publicly reported claims-based outcome measures).

Current Submission

Empiric Validity

Stewards of NQF-endorsed measures going through the re-endorsement process are required to demonstrate external validity testing at the time of maintenance review, or if this is not possible, justify the use of face validity

only. To address this requirement, we compared hospital performance on CMS's 30-day CABG mortality measure against two quality measures and one structural measure:

- The Society of Thoracic Surgeons (STS) CABG Composite Star Rating
- The mortality group score of CMS's Overall Hospital Star Ratings
- CABG procedural volume

Each comparator is described in more detail below.

STS CABG Composite

STS's CABG Composite Star Rating (ranging from 1-3 stars) is publicly reported online and calculated using a combination of 11 measures of isolated CABG quality divided into four broad domains (1. risk-adjusted 30-day mortality; 2. risk-adjusted major morbidity; 3. percentage of CABG procedures that use of internal mammary (or internal thoracic) artery for bypass grafting; and 4. prescription at discharge of beta-blockers, aspirin, and cholesterol-lowering medicines) (STS, 2022). The STS CABG Composite Star Rating calculation begins by assuming all providers are average and then determines statistically if there is at least a 99 percent probability that the performance of any specific provider is lower than average (one star) or higher than average (three star). These 11 individual measures and the overall composite measure methodology are all endorsed by the National Quality Forum. We compared the distribution of hospital-level RSMRs for CMS's 30-day CABG mortality level across each STS CABG Composite Star Rating category. We hypothesized that three-star hospitals would have lower average RSMRs (indicating lower mortality and higher quality) than one-star hospitals.

Mortality Group Score from CMS's Overall Hospital Star Ratings

CMS's Overall Hospital Star Rating assesses hospitals' overall performance (expressed on CMS' Care Compare) based on a weighted average of group scores from different domains of quality (mortality, readmissions, safety, patient experience, timely and effective care). The mortality group is comprised of the mortality measures that are publicly reported on Care Compare and in the updated methodology, is now calculated as a simple average of the performance on the individual mortality measures. For the validity testing presented in this testing form, we used mortality group scores from Medicare FFS hospitals from January 2020 and calculated the correlation with and without the CABG measure in the group score. The full methodology for the Overall Hospital Star Rating can be found at https://qualitynet.cms.gov/files/603966dda413b400224ddf50?filename=Star_Rtngs_CompMthdlgy_v4.1.pdf

We hypothesized that better performance on the 30-day CABG mortality measure (lower scores) would be associated with better performance (higher score) for the Mortality Group Score (moderate negative association). To compare performance, we calculated Pearson's Correlation Coefficient, and also plotted the relationship in a box plot.

Volume

There is some evidence, including a recent systematic review, of a weak but well-established hospital volume-outcome relationship for CABG procedures (Post et. al, 2010; Shahian & Normand, 2008; Birkmeyer et al., 2006). Therefore, we explored the relationship between 30-day CABG measure scores within deciles of CABG procedural volume. We expect that average RSMRs would decrease with increasing deciles of procedural volume, particularly at higher volumes.

References:

Birkmeyer JD, Dimick JB, Staiger DO. Operative mortality and procedure volume as predictors of subsequent hospital performance. Ann Surg. 2006 Mar;243(3):411-7.

Bratzler DW, Normand SL, Wang Y, et al. An administrative claims model for profiling hospital 30-day mortality rates for pneumonia patients. PLoS One 2011;6(4):e17401.

Keenan PS, Normand SL, Lin Z, et al. An administrative claims measure suitable for profiling hospital performance on the basis of 30-day all-cause readmission rates among patients with heart failure. Circulation 2008;1(1):29-37.

Krumholz HM, Brindis RG,Brush JE, et al. Standards for Statistical Models Used for Public Reporting of Health Outcomes: An American Heart Association Scientific Statement From the Quality of Care and Outcomes Research

Interdisciplinary Writing Group: Cosponsored by the Council on Epidemiology and Prevention and the Stroke Council Endorsed by the American College of Cardiology Foundation. Circulation. January 24, 2006 2006;113(3):456-462.

Krumholz HM, Wang Y, Mattera JA, et al. An administrative claims model suitable for profiling hospital performance based on 30-day mortality rates among patients with an acute myocardial infarction. Circulation 2006;113(13):1683-92.

Krumholz HM, Wang Y, Mattera JA, et al. An administrative claims model suitable for profiling hospital performance based on 30-day mortality rates among patients with heart failure. Circulation 2006;113:1693-1701.

National Quality Forum. National voluntary consensus standards for patient outcomes, first report for phases 1 and 2: A consensus report http://www.qualityforum.org/projects/Patient_Outcome_Measures_Phases1-2.aspx Accessed August 19, 2010.

Post, P. N., Kuijpers, M., Ebels, T., & Zijlstra, F. (2010). The relation between volume and outcome of coronary interventions: a systematic review and meta-analysis. European Heart Journal, 31(16), 1985–1992. https://doi.org/10.1093/eurheartj/ehq151

Shahian DM, He X, O'Brien S, et al. Development of a Clinical Registry-Based 30-Day Readmission Measure for Coronary Artery Bypass Grafting Surgery. Circulation 2014; DOI: 0.1161/CIRCULATIONAHA.113.007541. Published online before print June 10, 2014

Shahian DM, Normand SLT. Low-volume coronary artery bypass surgery: Measuring and optimizing performance. The Journal of Thoracic and Cardiovascular Surgery. 2008;135(6):1202-1209. doi:10.1016/j.jtcvs.2007.12.037

Suter L, Wang C, Araas M, et al. Hospital-Level 30-Day All-Cause Unplanned Readmission Following Coronary Artery Bypass Graft Surgery (CABG): Updated Measure Methodology Report. 2014; http://www.qualitynet.org/dcs/BlobServer?blobkey=id&blobnocache=true&blobwhere=1228890352615&blobheader=multipart%2Foctet-stream&blobheadername1=Content-Disposition&blobheadervalue1=attachment%3Bfilename%3DRdmsn_CABG_MeasMethd_Rpt_060314.pdf&blobcol=urldata&blobtable=MungoBlobs Accessed November 4, 2015.

The Society of Thoracic Surgeons. STS Public Reporting Online. CABG Overall Composite Score. 2022. Available at: https://publicreporting.sts.org/cabg-composite-score, accessed June 15, 2022.

Xian Y, Fonarow GC, Reeves MJ, et al. Data quality in the American Heart Association Get With The Guidelines-Stroke (GWTG-Stroke): Results from a National Data Validation Audit. American Heart Journal. 2012;163(3):392-398.e391. http://www.ahjonline.com/article/S0002-8703%2811%2900894-5/abstract

[Response Ends]

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

Examples may include correlations or t-test results.

[Response Begins]

Previous Submission

Validity as Assessed by External Groups

Fourteen TEP members responded to the survey question as follows: Strongly Disagreed (1), Moderately Disagreed (1), Somewhat Disagreed (1), Somewhat Agreed (1), Moderately Agreed (8), and Strongly Agreed (2). Hence, 79% of TEP members agreed (71% moderately or strongly agreed) that the measure will provide an accurate reflection of quality.

Validation of Administrative Risk Adjustment Model

The validation of the administrative risk model demonstrated similar distributions in hospital RSMRs for the claims-based and clinical-based models, although the claims-based model showed a narrower range of outcome rates.

The C-statistics for the two models were similar: 0.74 for the claims-based model and 0.75 for the clinical-based model. Overall agreement between hospital performance categorization between the claims-based and clinical-based models was 94.3% (33 of 35 hospitals had concordant performance categorization) and the correlation was 0.90 (weighted Spearman correlation). The clinical-based model identified two worse-performing outlier hospitals, while the claims-based model identified none; neither model identified any better-performing outliers in the matched sample.

Full results of the validation study can be found in the Appendix of the CABG Mortality Measure Methodology Report.

Current Submission

As described above, we assessed empiric validity by comparing performance on CMS's 30-day CABG mortality measure to two quality measures and one structural measure:

- The Society of Thoracic Surgeons (STS) CABG Composite Star Rating
- The mortality group score of CMS's Overall Hospital Star Ratings
- CABG procedural volume

STS CABG Composite Star Rating

We found a stepwise trend of lower 90-day mortality with higher STS Composite Star Ratings (Figure 1). The median 30-day all-cause CABG mortality RSMR was lower (better performance) with each increasing star of the STS Composite Star Ratings (better performance). The overall correlation between CMS's 30-day CABG measure scores and the STS CABG Composite Star Rating was -0.382.

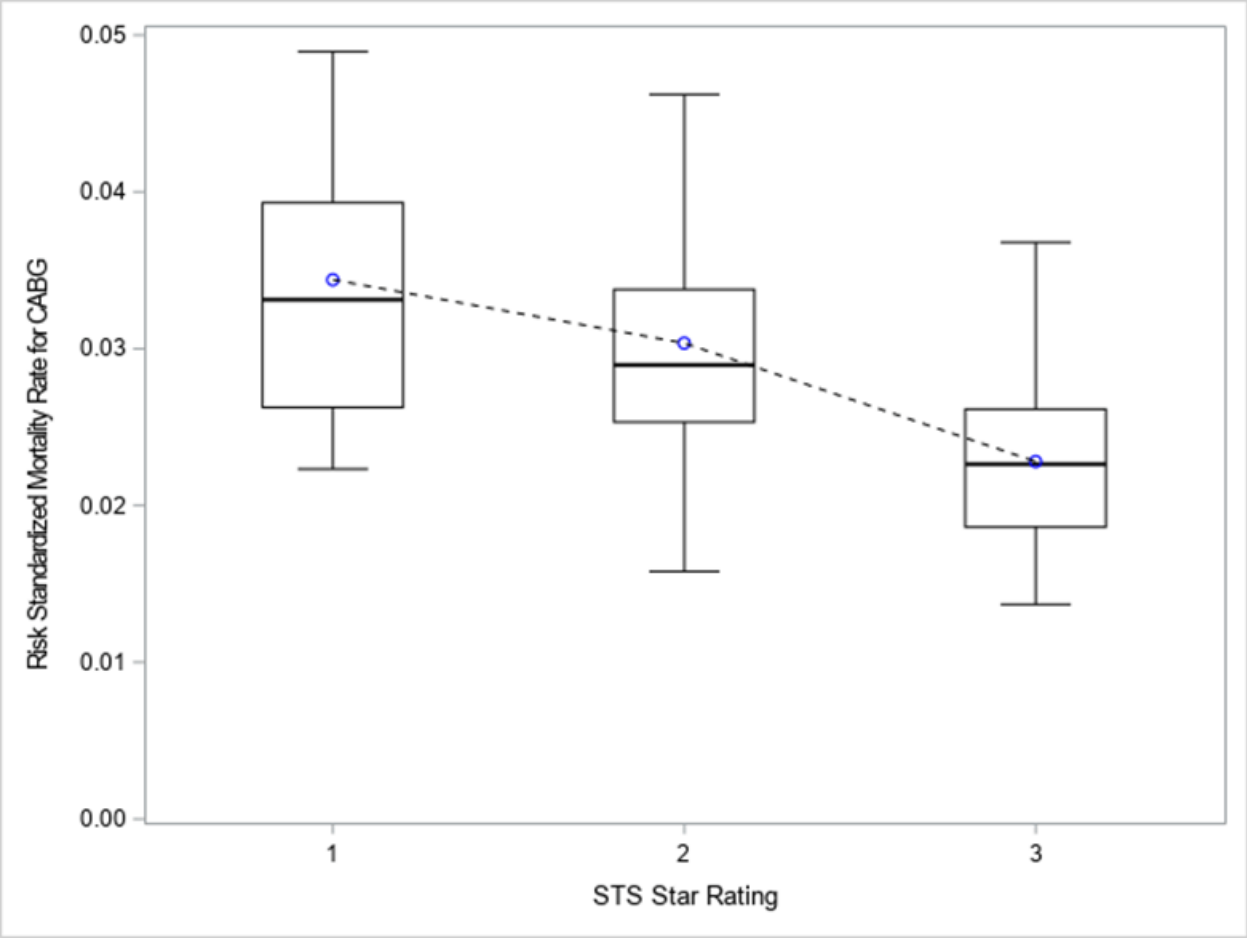


Figure 1: Empiric Validity Testing: CMS's 30-day CABG RSMR vs. STS CABG Composite Online Star Ratings

Mortality Group Score from CMS's Overall Hospital Star Ratings

30-day CABG RSMRs were moderately negatively correlated (-0.445) with CMS's Hospital Star Rating Mortality Group Scores; the relationship, while weaker (-0.276), was maintained even after removing the CABG mortality measure from the group (Table 3). Figure 2 shows a decreasing trend of RSMRs within quartiles of the Mortality Group Score.

Metric	Mortality group score WITH 30-day CABG measure in the group	Mortality group score WITHOUT 30-day CABG in the group
Pearson correlation coefficient for association between 30-day CABG mortality measure scores and Star Rating Mortality Group Score	-0.445 (p<.0001)	-0.276 (p<.0001)

Table 3: Association between 30-day CABG mortality measure scores and Mortality Group Scores

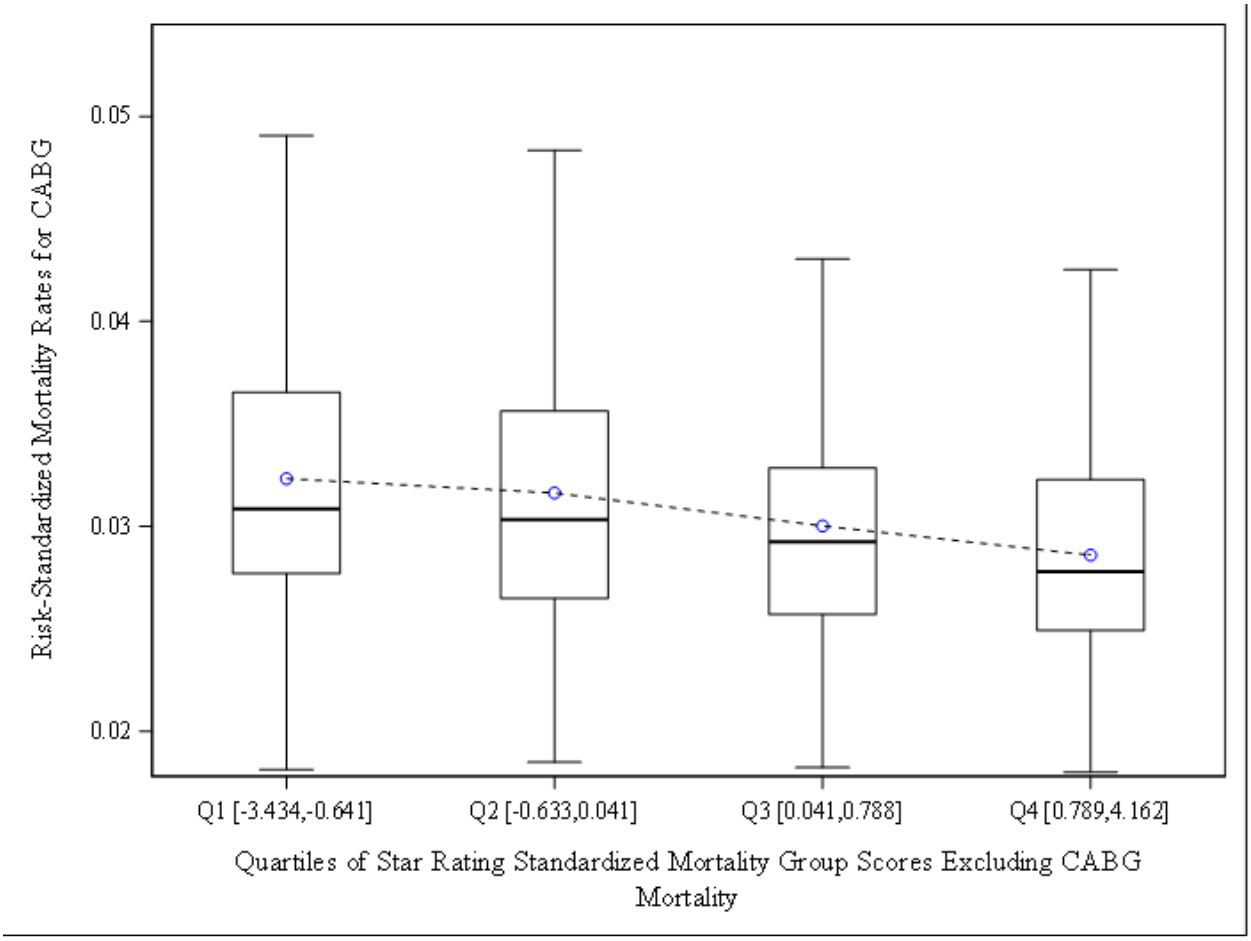


Figure 2: Distribution of 30-day CABG Mortality RSMRs within Quartiles of Hospital Star Rating Mortality Group Scores (removing 30-day CABG from the Mortality Group Score)

Procedural Volume

Among 1,002 hospitals with at least 25 cases, mean RSMRs declined with increasing CABG volume but at higher volumes, as expected. The overall correlation of CABG volume and 30-day CABG RSMR was -0.214 (p<0.05).

Volume Decile	Number of hospitals	Range of procedural volume	Mean 30-day CABG RSMRs
1	101	25-38	3.11
2	100	39-49	3.18
3	98	50-65	3.24
4	102	66-82	3.18
5	100	83-104	3.24
6	98	105-124	3.07
7	104	125-152	3.07
8	99	153-193	3.07
9	100	194-271	2.85
10	100	272-852	2.69

Table 4: Correlation of 30-day CABG mortality with admission volume as well as RSMR by declines of volume

References:

Curtis J, Drye E, Geary L, et al. Hospital 30-Day Percutaneous Coronary Intervention Mortality Measure: Center for Outcomes Research and Evaluation;2010.

Keenan PS, Normand SL, Lin Z, Drye EE, Bhat KR, Ross JS, Schuur JD, Stauffer BD, Bernheim SM, Epstein AJ, Wang Y-F, Herrin J, Chen J, Federer JJ, Mattera JA, Wang Y, Krumholz HM. An administrative claims measure suitable for profiling hospital performance on the basis of 30-day all-cause readmission rates among patients with heart failure. Circulation: Cardiovascular Quality and Outcomes. 2008 Sep;1(1):29-37.

Krumholz HM, Lin Z, Drye EE, Desai MM, Han LF, Rapp MT, Mattera JA, Normand SL. An administrative claims measure suitable for profiling hospital performance based on 30-day all-cause readmission rates among patients with acute myocardial infarction. Circulation: Cardiovascular Quality and Outcomes. 2011 Mar 1;4(2):243-52.

Lindenauer PK, Normand SL, Drye EE, Lin Z, Goodrich K, Desai MM, Bratzler DW, O´Donnell WJ, Metersky ML, Krumholz HM. Development, validation, and results of a measure of 30-day readmission following hospitalization for pneumonia. Journal of Hospital Medicine. 2011 Mar;6(3):142-50.

Shahian DM, Silverstein T, Lovett AF, Wolf RE, Normand SLT. Comparison of Clinical and Administrative Data Sources for Hospital Coronary Artery Bypass Graft Surgery Report Cards. Circulation. 2007; 115: 1518-1527.

[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]

Previous Submission

Validity as Assessed by External Groups

The results demonstrate TEP agreement with overall face validity of the measure as specified. Measure validity is also ensured through the processes employed during development, including regular expert and clinical input, and modeling methodologies with demonstrated validity in claims-based measures.

Validation of Administrative Risk Adjustment Model

Thorough evaluation adherent to nationally accepted standards for outcome measure development (Krumholz et al. 2008; Shahian et al. 2007) indicate that the model has similar discrimination and calibration to a New York

state-derived clinical risk model, although the relative discrimination was lower when a risk variable (shock), whose pre-operative status was unknown, was removed from the claims-based model. Although both the mortality rate and range of performance in the matched sample was less than that of US hospitals overall, the frequency and effect of risk variables was similar in the matched sample and national data. The models produce similar estimates of hospital performance. However, the claims-based model generally produced lower RSMR estimates compared with the clinical-based model among hospitals with higher estimated RSMRs, and higher RSMR estimates among those hospitals with lower RSMRs. Assuming that the clinical-based model is the gold standard (and does not over-estimate poor performing hospitals' RSMRs), our findings suggest that the claims-based model may underestimate poor performing hospitals' RSMRs and may be less likely to identify poor performance outliers compared with the clinical-based model. Similarly, the claims-based model may be less likely to identify hospitals with significantly better-than-average performance, although this validation study cannot assess this as the clinical-based model did not identify high performing outlier hospitals in the validation sample.

Current Submission

Empiric Validity Testing

The results of the measure score validation testing against the STS CABG Composite Star Ratings measure support the validity of the 30-day CABG measure; results demonstrate, as expected, a moderate correlation in the expected direction between the two measures. Empiric validity results comparing CMS Hospital Star Ratings Mortality Group scores showed the expected strength and direction of the relationship, and the association with volume was also as expected, with better performance on the 30-day CABG mortality measure at higher volumes.

The empiric validity results, taken together with the validation of the model against medical records, the face validity vote from the TEP as well as the face validity of the measure concept (mortality), support the overall validity of the 30-day CABG mortality measure.

References:

Krumholz HM, Keenan PS, Brush JE, Jr., et al. Standards for measures used for public reporting of efficiency in health care: a scientific statement from the American Heart Association Interdisciplinary Council on Quality of Care and Outcomes Research and the American College of Cardiology Foundation. *Circulation.* Oct 28 2008;118(18):1885-1893.

Shahian DM, Silverstein T, Lovett AF, Wolf RE, Normand SLT. Comparison of Clinical and Administrative Data Sources for Hospital Coronary Artery Bypass Graft Surgery Report Cards. Circulation.2007; 115: 1518-1527.

[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]

We demonstrate meaningful differences using two approaches:

1. Reporting the distribution of measure scores.
2. Reporting performance categories.

For public reporting of the measure, CMS characterizes the uncertainty associated with the RSMR by estimating the 95% interval estimate. This is similar to a 95% confidence interval but is calculated differently. If the RSMR's interval estimate does not include the national observed mortality rate (is lower or higher than the rate), then CMS is confident that the hospital's RSMR is different from the national rate, and describes the hospital on the Hospital Compare website as "better than the U.S. national rate" or "worse than the U.S. national rate." If the interval includes the national rate, then CMS describes the hospital's RSMR as "no different than the U.S. national rate" or

"the difference is uncertain." CMS does not classify performance for hospitals that have fewer than 25 cases in the three-year period.

[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]

Previous submission

Analyses of Medicare FFS data show substantial variation in RSMRs among hospitals. Using data from July 2013-June 2016 (**Dataset 1**), the median hospital RSMR was 3.1%, with a range of 1.3% to 7.4%. The interquartile range was 2.7%-3.7%.

Of 1,185 hospitals in the study cohort, 17 performed "Better than the National Rate," 1,004 performed "No Different from the National Rate," and 18 performed "Worse than the National Rate." 146 were classified as "Number of Cases Too Small" (fewer than 25) to reliably tell how well the hospital is performing.

Current Submission

Measure Score Distribution

Updated analyses of Medicare FFS data show substantial variation in RSMRs among hospitals. Using data from July 2016-June 2019 (Fall 2022 Endorsement Maintenance Dataset), the median hospital RSMR was 2.9%, with a range of 1.4% to 6.8%. The interquartile range was 2.6%-3.4%. Figure 3 shows a histogram of hospital performance on the 30-day CABG mortality measure.

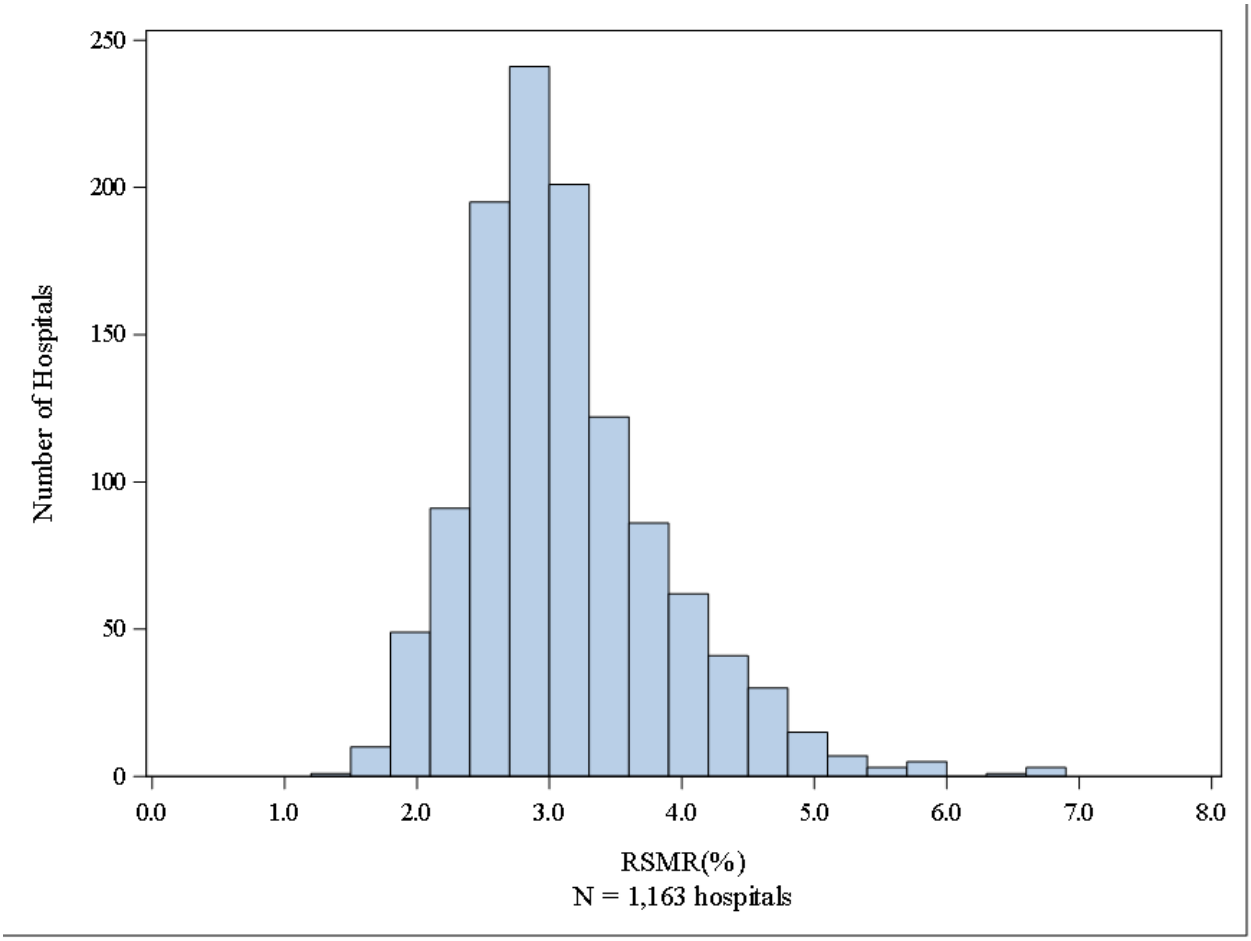


Figure 3: 30-day CABG Mortality: Histogram of hospital performance (Fall 2022 Endorsement Maintenance Dataset)

Performance categories

Of 1,163 hospitals in the study cohort, 13 performed "Better than the National Rate," 974 performed "No Different from the National Rate," and 15 performed "Worse than the National Rate." 161 were classified as "Number of Cases Too Small" (fewer than 25) to reliably tell how well the hospital is performing.

[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]

Previous Submission

The variation in rates suggests there are meaningful differences across hospitals in 30-day all-cause mortality following a qualifying CABG procedure.

Current Submission

With updated data, measure score distribution shows that there is still significant variation in measure scores. For example, the hospital with the best RSMR (1.4%) is performing 52% better than the average (median) performer and the hospital with the worst RSMR (6.8%) is performing 134% worse than the average (median) performer. (We note that the "average" performer refers to a facility with the same case and procedure mix performing at the average (median)).

[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]

The 30-day CABG Mortality measure used claims-based data for development and testing. There was no missing data in the claims-based development and testing data.

[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]

Not applicable.

[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]

Not applicable.

[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]

No, there is only one set of specifications for this measure

[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]

Yes, the measure uses exclusions.

[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]

Previous submission

All exclusions were determined by careful clinical review and have been made based on clinically relevant decisions and to ensure accurate calculation of the measure. To ascertain impact of exclusions on the cohort, we examined overall frequencies and proportions of the total cohort excluded for each exclusion criterion (**Dataset 1**). These exclusions are consistent with similar NQF-endorsed outcome measures. Rationales for the exclusions are detailed in field S.9 of the measure submission form (Denominator Exclusions Details).

Current submission

We updated the results showing the proportion and distribution of exclusions for this measure using the Fall 2022 Endorsement Maintenance Dataset.

[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]

Previous submission

Exclusion	N	%	Distribution across hospitals (N=1,039): Min, 25th, 50th, 75th percentile, max
1. Inconsistent or unknown vital status or other unreliable demographic data	1	<0.01%	(0, 0, 0, 0, 0.25)
2. Admissions for subsequent qualifying CABG procedures during the measurement period	88	0.06%	(0, 0, 0, 1.19, 3.45)
3. Discharged against medical advice (AMA)	46	0.03%	(0, 0, 0, 0, 3.70)

Table 5: Distribution of exclusions among hospitals with 25 or more admissions

Current Submission

Exclusion	N	%	Distribution across hospitals (N=1,002): Min, 25th, 50th, 75th percentile, max
1. Inconsistent or unknown vital status or other unreliable demographic data	2	0.00%	(0.00,0.00,0.00,0.00,1.11)
2. Admissions for subsequent qualifying CABG procedures during the measurement period	40	0.06%	(0, 0, 0, 1.19, 3.45)
3. Discharged against medical advice (AMA)	44	0.03%	(0.00,0.00,0.00,0.00,3.13)

Table 6: Distribution of exclusions among hospitals with 25 or more admissions

[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]

Previous Submission

Exclusion 1 is necessary for valid calculation of the measure. Patients with an inconsistent or unknown vital status or other unreliable demographic account for <0.01% of all index admissions excluded from the initial index cohort.

Exclusion 2 (admissions for subsequent qualifying CABG procedures during the measurement period) accounts for 0.06% of all index admissions excluded from the initial index cohort. This exclusion was applied to align with the Society of Thoracic Surgeons 30-day mortality measure. The experts believed that a second CABG procedure within 30 days of an initial procedure is most likely due to a complication of the initial CABG procedure or the peri-operative care the patient received, and as such, the care provided by the hospital performing the initial CABG procedure likely dominates mortality risk.

Exclusion 3 (patients who are discharged AMA) accounts for 0.03% of all index admissions excluded from the initial index cohort. This exclusion is needed for acceptability of the measure to hospitals, who do not have the opportunity to deliver full care and prepare the patient for discharge.

Current submission

Results from the Fall 2022 Endorsement Maintenance Dataset are similar to the previous submission. The rationale for the exclusions remain the same. In total, less than 1% of the cohort are excluded from the measure.

[Response Ends]

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

[Response Begins]

Statistical risk model with risk factors (specify number of risk factors)

[Statistical risk model with risk factors (specify number of risk factors) Please Explain]

24

[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]

See codes that define the risk model variables in the attached data dictionary. Final risk model variables and odds ratios are reported in section 2b.24.

The CABG surgery measure uses a hierarchical generalized linear model (HGLM) to estimate RSMRs for hospitals. This modeling approach accounts for the within-hospital correlation of the observed outcome and accommodates the assumption that underlying differences in quality across hospitals lead to systematic differences in outcomes.

In the CABG surgery measure, an HGLM model is estimated. Then for each hospital, a standardized mortality ratio (SMR) is calculated. The RSMR is calculated by multiplying the SMR for each hospital by the national observed mortality rate.

Hierarchical Generalized Linear Model

We fit an HGLM, which accounts for clustering of observations within hospitals. We assume the outcome has a known exponential family distribution and relates linearly to the covariates via a known link function, *h*. Specifically, we assume a binomial distribution and a logit link function. Further, we account for the clustering within hospitals by estimating a hospital-specific effect, □□, which we assume follows a normal distribution with a mean □ and variance □2, the between-hospital variance component. The following equation defines the HGLM:

$$h\left(\Pr\left(Y_{ij}=1\middle|\boldsymbol{Z}_{ij}-\omega_i\right)\right)=log\left(\frac{\Pr\left(Y_{ij}=1|\boldsymbol{Z}_{ij}-\omega_i\right)}{1-\Pr\left(Y_{ij}=1|\boldsymbol{Z}_{ij}-\omega_i\right)}\right)=\alpha_i+\boldsymbol{\beta}\boldsymbol{Z}_{ij} \qquad (1)$$

$$\text{where } \alpha_i=\mu+\omega_i;\ \omega_i\sim N(0-\tau^2)$$

$$i=1,\ldots,I;\ j=1,\ldots,n_i$$

Equation 1

where □□□ denotes the outcome (equal to 1 if the patient dies within 30 days, 0 otherwise) for the *j-th* patient at the *i-th* hospital; □□□□=(□□□1− □□□2− …− □□□□)is a set of *p* patient-specific covariates derived from the data; and *I* denotes the total number of hospitals and □□ denotes the number of index admissions at hospital *i*. The hospital-specific intercept of the *i-th* hospital, □□, defined above, comprises □, the adjusted average intercept over all hospitals in the sample, and □□, the hospital-specific intercept deviation from □.

We estimate the HGLM using the SAS software system (GLIMMIX procedure).

[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]

Published literature

[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]

Our goal was to develop a parsimonious model that included clinically relevant variables associated with isolated CABG mortality. The candidate variables for the model were derived from: the index admission, with comorbidities identified from the index admission secondary diagnoses (excluding potential complications), 12-month pre-index inpatient Part A data, outpatient hospital data, and Part B physician data.

For administrative model development, we started with 189 Condition Categories (CCs) which are part of CMS's Hierarchical Condition Categories. The Hierarchical Condition Category (HCC) system groups the ICD-9-CM codes into larger groups that are used in models to predict medical care utilization, mortality, or other related measures. CCs are clinically relevant diagnostic groups of the more than 15,000 ICD-9 codes (Pope et al. 2001).

To select candidate variables, a team of clinicians reviewed all 189 CCs and excluded those that were not relevant to the Medicare population or that were not clinically relevant to the mortality outcome (e.g., attention deficit disorder, female infertility). Clinically relevant CCs were selected as candidate variables and some of those CCs were then combined into clinically coherent CC groupings. Other candidate variables included age, gender, and cardiogenic shock. Gender was included in risk adjustment due to the fact that women have smaller caliber vessels and thus represent more technically challenging CABG procedures compared to men (O'Connor 1996).

To inform final variable selection, a modified approach to stepwise logistic regression was performed. The development sample was used to create 1,000 "bootstrap" samples. For each sample, we ran a logistic stepwise regression that included the candidate variables. The results were summarized to show the percentage of times that each of the candidate variables was significantly associated with mortality ($p<0.001$) in each of the 1,000 repeated samples (e.g., 90 percent would mean that the candidate variable was selected as significant at $p<0.001$ in 90 percent of the estimations). We also assessed the direction and magnitude of the regression coefficients.

The clinical team reviewed these results and decided to retain the majority of risk adjustment variables above a 70% cutoff, because they demonstrated a relatively strong and stable association with risk for death and were clinically relevant. Additionally, specific variables with particular clinical relevance to the risk of death were forced into the model (regardless of percent selection) to ensure appropriate risk adjustment for CABG. These included:

1. Clinical variables associated with CABG:
    i. History of Prior CABG or Valve Surgery
2. Markers for end of life/frailty:
    i. Decubitus Ulcer or Chronic Skin Ulcer
    ii. Dementia or Other Specified Brain Disorders
    iii. Metastatic Cancer and Acute Leukemia
    iv. Protein-calorie Malnutrition
    v. Hemiplegia, Paraplegia, Paralysis, Functional disability
    vi. Stroke
3. Diagnoses with potential asymmetry among hospitals that would impact the validity of the model:
    i. Lung, Upper Digestive Tract, and Other Severe Cancers
    ii. Lymphatic, Head and Neck, Brain, and Other Major Cancers; Breast, Prostate, Colorectal and Other Cancers and Tumors; Other Respiratory and heart Neoplasms
    iii. Other Digestive and Urinary Neoplasms

This resulted in a final risk-adjustment model that included 24 variables.

Previous Submission

Social Risk Factors

We selected variables representing social risk factors such as socioeconomic status for examination based on a review of literature, conceptual pathways, and feasibility. In Section 1.8, we describe the variables that we

considered and analyzed based on this review. Below we describe the pathways by which social risk factors may influence 30-day mortality.

Our conceptualization of the pathways by which patient social risk factors affect 30-day mortality is informed by the literature.

Literature Review of Social Risk Variables and Mortality after a CABG Procedure

To examine the relationship between social risk factors and hospital 30-day, all-cause, RSMR following CABG surgery, a literature search was performed with the following exclusion criteria: international studies, articles published more than 10 years ago, articles without primary data, articles using Veterans Affairs databases as the primary data source, and articles not explicitly focused on social risk factors such as SES and CABG mortality. Studies are limited, and those that have been conducted have mixed results.

Causal Pathways for Social Risk Variable Selection

Although some recent literature evaluates the relationship between patient social risk factor such as SES and the mortality outcome, few studies directly address causal pathways or examine the role of the hospital in these pathways (see, for example, Chang et al 2007; Gopaldas et al 2009; Kim et al 2007; LaPar 2010; 2012). Moreover, the current literature examines a wide range of conditions and risk variables with no clear consensus on which risk factors demonstrate the strongest relationship with mortality. The social risk factors that have been examined in the literature can be categorized into three domains: (1) patient-level variables, (2) neighborhood/community-level variables, and (3) hospital-level variables. Patient-level variables describe characteristics of individual patients and include the patient's income or education level (Eapen et al., 2015). Neighborhood/community-level variables use information from sources such as the American Community Survey as either a proxy for individual patient-level data or to measure environmental factors. Studies using these variables use one dimensional measures such as median household income or composite measures such as the AHRQ-validated SES index score (Blum et al., 2014). Hospital-level variables measure attributes of the hospital which may be related to patient risk. Examples of hospital-level variables used in studies are ZIP code characteristics aggregated to the hospital level or the proportion of Medicaid patients served in the hospital (Gilman et al., 2014; Joynt and Jha, 2013).

The conceptual relationship, or potential causal pathways by which these possible social risk factors influence the risk of mortality following an acute illness or major surgery, like the factors themselves, are varied and complex. There are at least four potential pathways that are important to consider.

1. **Relationship of social risk factors such as SES to health at admission**. Patients who have lower income/education/literacy or unstable housing may have a worse general health status and may present for their hospitalization or procedure with a greater severity of underlying illness. These social risk factors, which are characterized by patient-level or neighborhood/community-level (as proxy for patient-level) variables, may contribute to worse health status at admission due to competing priorities (restrictions based on job, lack of childcare), lack of access to care (geographic, cultural, or financial), or lack of health insurance. Given that these risk factors all lead to worse general health status, this causal pathway should be largely accounted for by current clinical risk-adjustment**.**

2. **Use of low-quality hospitals**. Patients of lower income, lower education, or unstable housing have been shown not to have equitable access to high quality facilities because such facilities are less likely to be found in geographic areas with large populations of poor patients. Thus, patients with low income are more likely to be seen in lower quality hospitals, which can explain increased risk of mortality following hospitalization.

3. **Differential care within a hospital**. The third major pathway by which social risk factors may contribute to mortality risk is that patients may not receive equivalent care within a facility. For example, patients with social risk factors such as lower education may require differentiated care (e.g. provision of lower literacy information – that they do not receive).

4. **Influence of social risk factors on mortality risk outside of hospital quality and health status**. Some social risk factors, such as income or wealth, may affect the likelihood of mortality without directly affecting health status at admission or the quality of care received during the hospital stay. For instance, while a hospital may make appropriate care decisions and provide tailored care and education, a lower-income patient may have a worse outcome post-discharge due to competing economic priorities or a lack of access to care outside of the hospital.

These proposed pathways are complex to distinguish analytically. They also have different implications on the decision to risk adjust or not. We, therefore, first assessed if there was evidence of a meaningful effect on the risk model to warrant efforts to distinguish among these pathways.

Based on this model and the considerations outlined in Section 1.8, the following social risk variables were considered:

- Dual eligible status
- AHRQ SES index

We assessed the relationship between the SES variables with the outcome and examined the incremental effect in a multivariable model. For this measure, we also examined the extent to which the addition of any one of these variables improved model performance or changed hospital results. Given no meaningful improvement in the risk-model or change in performance scores we did not further seek to distinguish the causal pathways for these measures.

Current Submission

We have updated our literature search and provided an additional conceptual model for this Fall 2022 endorsement maintenance submission. Our literature search included articles published in the last ten years that addressed outcomes for patients undergoing CABG procedures, with a focus on patients with social risk factors.

There are well-established racial and sex-based differences in outcomes (mortality) for patients undergoing CABG procedures: Black patients, as well as women, have worse outcomes when compared with their white, male, counterparts (Mehta et al., 2016; Keeling et al., 2017). While there has been some improvement in these disparities over time, two recent articles affirm these continued disparities, one using claims data for older Americans (Angraal et al., 2018), and one using registry data from the Society for Thoracic Surgeons (STS) database (Enumah et al., 2020). The STS study, in addition, explored the underlying causes of the disparities. In this study, the authors found that although Black patients had a higher proportion of comorbidities, they still had worse outcomes compared with white patients even when controlling for those comorbidities. Furthermore, Black patients still had worse outcomes even after additionally controlling for socioeconomic status, insurance, sex, and case status. Females also had worse outcomes after controlling for all known variables. The study authors noted the potential role for hospital, community and local factors and pointed to studies that have found that a major component of disparate outcomes is the fact that Black and minority patients are admitted in higher proportions at low-quality hospitals (Khera et al., 2015).

Khera et al., previously published a conceptual model for the impact of risk factors on outcomes following CABG procedures, shown in Figure 4. The model divides variables into categories that represent biological differences, baseline health status, socioeconomic factors, cultural differences, as well as hospital and provider quality. Based on this conceptual model, published literature, and the availability of variables that can be linked to claims data, we have examined the following risk factors for the 30-day CABG mortality measure (each variable is discussed in more detail in section 2a.08):

- Dual eligibility – a patient-level variable and proxy for low income
- Low AHRQ SES – a 9-digit-zip-code-level variable, based on the AHRQ SES Index which is a multicomponent index and includes income, education, housing, and employment data
- Black race – a race variable within Medicare administrative data

We did not include testing for the sex variable because the current risk model already adjusts for patients' sex.

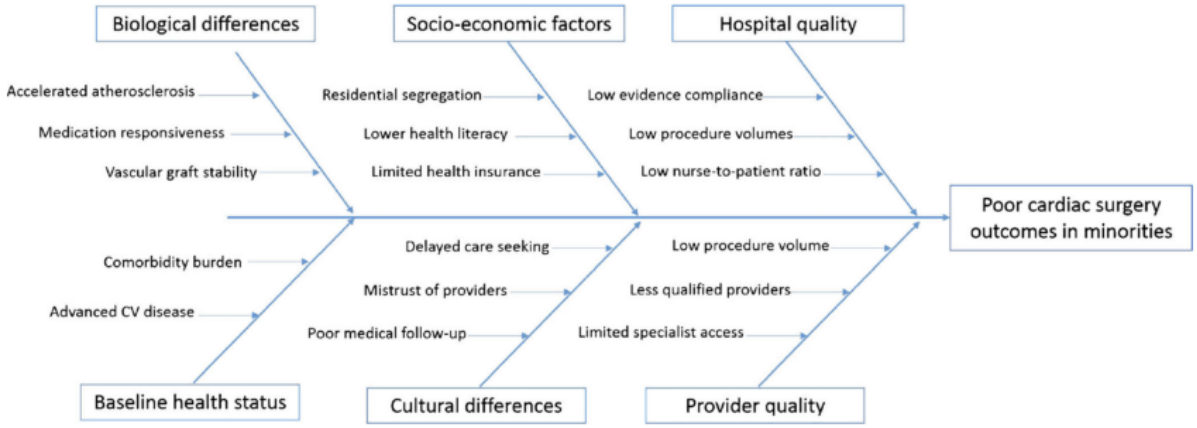


Fig. 1.
Root cause analysis of higher mortality in black and other minority patients after cardiac surgery

Figure 4: From Kehra et al., 2015

References:

Angraal S, Khera R, Wang Y, Lu Y, Jean R, Dreyer RP, Geirsson A, Desai NR, Krumholz HM. Sex and Race Differences in the Utilization and Outcomes of Coronary Artery Bypass Grafting Among Medicare Beneficiaries, 1999-2014. J Am Heart Assoc. 2018 Jul 12;7(14):e009014.

Blum, A. B., N. N. Egorova, E. A. Sosunov, A. C. Gelijns, E. DuPree, A. J. Moskowitz, A. D. Federman, D. D. Ascheim and S. Keyhani. Impact of Socioeconomic Status Measures on Hospital Profiling in New York City. Circ Cardiovasc Qual Outcomes 7, no. 3 (2014): 391-7.

Calvillo-King L, Arnold D, Eubank KJ, et al. Impact of social factors on risk of readmission or mortality in pneumonia and heart failure: systematic review. Journal of general internal medicine. 2013;28(2):269-282.

Chang W-C, Kaul P, Westerhout C M, Graham M. M., Armstrong Paul W., Effects of Socioeconomic Status on Mortality after Acute Myocardial Infarction. The American Journal of Medicine. 2007; 120(1): 33-39

Eapen ZJ, McCoy LA, Fonarow GC, Yancy CW, Miranda ML, Peterson ED, Califf RM, Hernandez AF. Utility of socioeconomic status in predicting 30-day outcomes after heart failure hospitalization. Circ Heart Fail. May 2015; 8(3):473-80.

Enumah ZO, Canner JK, Alejo D, Warren DS, Zhou X, Yenokyan G, Matthew T, Lawton JS, Higgins RSD. Persistent Racial and Sex Disparities in Outcomes After Coronary Artery Bypass Surgery: A Retrospective Clinical Registry Review in the Drug-eluting Stent Era. Ann Surg. 2020 Oct;272(4):660-667.

Foraker, R. E., K. M. Rose, C. M. Suchindran, P. P. Chang, A. M. McNeill and W. D. Rosamond. Socioeconomic Status, Medicaid Coverage, Clinical Comorbidity, and Rehospitalization or Death after an Incident Heart Failure Hospitalization: Atherosclerosis Risk in Communities Cohort (1987 to 2004). Circ Heart Fail 4, no. 3 (2011): 308-16.

Gopaldas R R, Chu D., Predictors of surgical mortality and discharge status after coronary artery bypass grafting in patients 80 years and older. The American Journal of Surgery. 2009; 198(5): 633-638

Gilman M, Adams EK, Hockenberry JM, Wilson IB, Milstein AS, Becker ER. California safety-net hospitals likely to be penalized by ACA value, readmission, and meaningful-use programs. Health Affairs (Millwood). Aug 2014; 33(8):1314-22.

Joynt KE, Jha AK. Characteristics of hospitals receiving penalties under the Hospital Readmissions Reduction Program. JAMA. Jan 23 2013; 309(4):342-3.

Keeling WB, Binongo J, Halkos ME, Leshnower BG, Nguyen DQ, Chen EP, Sarin EL, Miller JS, Macheers S, Lattouf OM, Guyton RA, Thourani VH. The Racial Paradox in Multiarterial Conduit Utilization for Coronary Artery Bypass Grafting. Ann Thorac Surg. 2017 Apr;103(4):1214-1221.

Khera R, Vaughan-Sarrazin M, Rosenthal GE, Girotra S. Racial disparities in outcomes after cardiac surgery: the role of hospital quality. Curr Cardiol Rep. 2015 May;17(5):29.

Kind, A. J., S. Jencks, J. Brock, M. Yu, C. Bartels, W. Ehlenbach, C. Greenberg and M. Smith. Neighborhood Socioeconomic Disadvantage and 30-Day Rehospitalization: A Retrospective Cohort Study. Ann Intern Med 161, no. 11 (2014): 765-74.

Kim C, Diez A V, Diez Roux T, Hofer P, Nallamothu B K, Bernstein S J, Rogers M, Area socioeconomic status and mortality after coronary artery bypass graft surgery: The role of hospital volume. Clinical Investigation Outcomes, Health Policy, and Managed Care. 2007; 154(2): 385-390

Krumholz HM, Brindis RG, Brush JE, et al. 2006. Standards for Statistical Models Used for Public Reporting of Health Outcomes: An American Heart Association Scientific Statement From the Quality of Care and Outcomes Research Interdisciplinary Writing Group: Cosponsored by the Council on Epidemiology and Prevention and the Stroke Council Endorsed by the American College of Cardiology Foundation. Circulation 113: 456-462.

LaPar D J, Bhamidipati C M, et al. Primary Payer Status Affects Mortality for Major Surgical Operations. Annals of Surgery. 2010; 252(3): 544-551

LaPar D J, Stukenborg G J, et al Primary Payer Status Is Associated With Mortality and Resource Utilization for Coronary Artery Bypass Grafting. Circulation. 2012; 126:132-139

Mehta RH, Shahian DM, Sheng S, O'Brien SM, Edwards FH, Jacobs JP, Peterson ED; Society of Thoracic Surgery National Adult Cardiac Surgery Database; Duke Clinical Research Institute. Association of Hospital and Physician Characteristics and Care Processes With Racial Disparities in Procedural Outcomes Among Contemporary Patients Undergoing Coronary Artery Bypass Grafting Surgery. Circulation. 2016 Jan 12;133(2):124-30.

Normand S-LT, Shahian DM. 2007. Statistical and Clinical Aspects of Hospital Outcomes Profiling. Stat Sci 22 (2): 206-226. Reames BN, Birkmeyer NJ, Dimick JB, Ghaferi AA. Socioeconomic disparities in mortality after cancer surgery: failure to rescue. JAMA surgery 2014; 149:475-81.

O'Connor NJ, Morton JR, Birkmeyer JD, Olmstead EM, O'Connor GT. Effect of coronary artery diameter in patients undergoing coronary bypass surgery. Northern New England Cardiovascular Disease Study Group*. Circulation. 1996;*93(4):652-655.

Pope G, Ellis R, Ash A, et al. Principal Inpatient Diagnostic Cost Group Models for Medicare Risk Adjustment. Health Care Financing Review. 2000;21(3):26.

Regalbuto R, Maurer MS, Chapel D, Mendez J, Shaffer JA. Joint Commission requirements for discharge instructions in patients with heart failure: is understanding important for preventing readmissions? Journal of cardiac failure. 2014;20(9):641-649.

[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]

The tables below show the final variables in the model in the testing dataset with associated odds ratios (OR) and 95 percent confidence intervals (CI).

Previous Submission

Variable	07/2013-06/2016, OR (95% CI)
Age minus 65 (years above 65, continuous)	1.06 (1.06 - 1.07)
Male	0.69 (0.64 - 0.74)
Cardiogenic shock	7.20 (6.68 - 7.75)
Coronary atherosclerosis	1.18 (1.06 - 1.33)
History of coronary artery bypass graft (CABG) or valve surgery	1.41 (1.24 - 1.60)
Cancer; metastatic cancer and acute leukemia (CC 8-14)	0.92 (0.84 - 1.00)
Protein-calorie malnutrition (CC 21)	1.72 (1.55 - 1.91)
Morbid obesity; other endocrine/metabolic/nutritional disorders (CC 22, 25-26)	0.73 (0.66 - 0.82)
Liver or biliary disease (CC 27-32)	1.50 (1.35 - 1.67)
Other gastrointestinal disorders (CC 38)	0.77 (0.72 - 0.82)
Dementia or other specified brain disorders (CC 51-53)	1.29 (1.16 - 1.45)
Hemiplegia, paraplegia, paralysis, functional disability (CC 70-74, 103-104, 189-190)	1.29 (1.10 - 1.52)
Congestive heart failure (CC 85)	1.17 (1.08 - 1.27)
Acute myocardial infarction (CC 86)	1.20 (1.11 - 1.29)
Unstable angina and other acute ischemic heart disease (CC 87)	0.87 (0.81 - 0.93)
Angina; old myocardial infarction (CC 88 plus ICD-10-CM code I25.2, for discharges on or after October 1, 2015; CC 88 plus ICD-9-CM code 412, for discharges prior to October 1, 2015)	0.87 (0.81 - 0.93)
Hypertension (CC 95)	0.81 (0.74 - 0.89)
Stroke (CC 99-100)	1.06 (0.92 - 1.22)
Vascular or circulatory disease (CC 106-109)	1.16 (1.08 - 1.24)
Chronic obstructive pulmonary disease (COPD) (CC 111)	1.38 (1.29 - 1.48)
Pneumonia (CC 114-116)	1.32 (1.21 - 1.43)
Dialysis status (CC 134)	1.92 (1.66 - 2.23)
Renal failure (CC 135-140)	1.39 (1.30 - 1.49)
Decubitus ulcer or chronic skin ulcer (CC 157-161)	1.11 (0.97 - 1.28)

Table 7: Risk model variable odds ratios

Current Submission

Variable	07/2016-06/2019 OR (95% CI)
Age minus 65 (years above 65, continuous)	1.06 (1.05-1.07)
Male	0.63 (0.59-0.67)
Cardiogenic shock	3.65 (3.25-4.09)
Coronary atherosclerosis	1.12 (0.99-1.25)
History of coronary artery bypass graft (CABG) or valve surgery	1.49 (1.34-1.65)
Cancer; metastatic cancer and acute leukemia (CC 8-14)	0.93 (0.86-1.01)

Variable	07/2016-06/2019 OR (95% CI)
Protein-calorie malnutrition (CC 21)	2.09 (1.88-2.32)
Morbid obesity; other endocrine/metabolic/nutritional disorders (CC 22, 25-26)	0.62 (0.55-0.70)
Liver or biliary disease (CC 27-32)	1.58 (1.43-1.74)
Other gastrointestinal disorders (CC 38)	0.75 (0.70-0.80)
Dementia or other specified brain disorders (CC 51-53)	1.16 (1.03-1.30)
Hemiplegia, paraplegia, paralysis, functional disability (CC 70-74, 103-104, 189-190)	0.99 (0.83-1.17)
Congestive heart failure (CC 85)	1.20 (1.11-1.30)
Acute myocardial infarction (CC 86)	1.36 (1.26-1.47)
Unstable angina and other acute ischemic heart disease (CC 87)	0.95 (0.89-1.02)
Angina; old myocardial infarction (CC 88 plus ICD-10-CM code I25.2, for discharges on or after October 1, 2015; CC 88 plus ICD-9-CM code 412, for discharges prior to October 1, 2015)	0.76 (0.71-0.82)
Hypertension (CC 95)	0.73 (0.67-0.80)
Stroke (CC 99-100)	0.98 (0.84-1.14)
Vascular or circulatory disease (CC 106-109)	1.13 (1.06-1.22)
Chronic obstructive pulmonary disease (COPD) (CC 111)	1.44 (1.34-1.55)
Pneumonia (CC 114-116)	1.47 (1.35-1.60)
Dialysis status (CC 134)	1.73 (1.49-2.01)
Renal failure (CC 135-140)	1.51 (1.41-1.62)
Decubitus ulcer or chronic skin ulcer (CC 157-161)	1.01 (0.87-1.18)

Table 8: Risk model variable odds ratios

[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]

Analysis #1: Variation in prevalence of the factor across measured entities

Previous submission

The prevalence of SES factors in the CABG cohort varies across measured entities. The median percentage of dual eligible patients is 8.4% (interquartile range [IQR]: 5.6% – 13.4%). The median percentage of patients with an AHRQ SES Index score equal to or below 46.0 is 16.6% (IQR: 8.8% – 26.8%).

Current submission

The prevalence of SES factors in the CABG cohort varies across measured entities. The median percentage of dual eligible patients is 4.7% (interquartile range [IQR]: 2.3% - 8.7%). The median percentage of patients with an AHRQ SES Index score equal to or below 46.0 is 14.7% (IQR: 7.7% - 26.3%). The median percentage of Black patients is 2.7% (IQR: 0.0% - 7.4%).

Social Risk Factors	Median of the Hospital Prevalence of the social risk factor (IQR)
Dual Eligibility	**4.70% (2.30-8.70%)**
Low AHRQ SES	**14.7% (7.70-26.3%)**
Race (Black)	**2.70% (0.00-7.40%)**

Table 9: Facility-level distribution of social risk factors in the CABG mortality cohort

Analysis #2: Empirical association with the outcome (univariate)

Previous submission

The patient-level observed CABG mortality rate is higher for dual eligible patients, 4.68%, compared with 3.03% for all other patients. Similarly, the mortality rate for patients with an AHRQ SES Index score equal to or below 42.6 was 3.96% compared with 3.00% for patients with an AHRQ SES Index score above 42.6.

Current submission

The patient-level observed 30-day CABG mortality rate is higher for dual eligible patients, 4.0%, compared with 2.9% for all other patients. Similarly, the mortality rate for patients with an AHRQ SES Index score equal to or below 46 was higher; 3.9% compared with 2.8% for patients with an AHRQ SES Index score above 46. Observed 30-day mortality for Black patients was also higher; 3.8%, compared with 2.9% for all other patients.

Social Risk Factor	Mean Observed 30-day Mortality Rate
Dual Eligibility (DE)	**4.0%**
Non-DE	**2.9%**
Low AHRQ SES	**3.9%**
Non Low AHRQ SES	**2.8%**
Race (Black)	**3.8%**
All other race variables	**2.9%**

Table 10: Mean observed outcomes for patients with each social risk factor compared with patients without the social risk factor

Analysis #3: Incremental effect of SES variables in a multivariable model

Previous submission

We then examined the strength and significance of the SES variables in the context of a multivariable model. Consistent with the above findings, when we include any of these variables in a multivariate model that includes all of the claims-based clinical variables, the effect size of each of these variables is significant, but lower, than the coefficient for the bivariate association (the parameter estimate decreased from 1.57 to 1.23 for dual eligibility, from 1.34 to 1.23 for low AHRQ SES Index).

Current submission

When we include any of these variables in a multivariate model that includes all of the claims-based clinical variables, the effect size is lower than the coefficient for the bivariate association (the parameter estimate decreased from 1.38 to 1.0 for dual eligibility, from 1.30 to 1.27 for the low AHRQ SES Index, and from 1.26 to 1.02 for the Black race variable. In the multivariate model, only the low AHRQ SES Index variable remains significant (Table 11).

Social risk factor	Odds ratio for bivariate relationship	p-value	Odds ratio for multivariable relationship	p-value
Dual Eligible	1.38	<.0001	1.01	0.8662
Low AHRQ SES Index	1.3	<.0001	1.18	0.0001
Race (Black)	1.26	0.0005	1.01	0.858

Table 11: Odds ratios for 30-day mortality, bivariate and multivariate models

Analysis #4: Impact of social risk factors model performance

Previous submission

To further understand the relative importance of these risk-factors in the measure, we compared model performance with and without the addition of each social risk variable. Results show that the c-statistic is unchanged with the addition of any of these variables into the model: The c-statistic of the original model is 0.779; the c-statistic of the original model with the dual eligible variable added is 0.779; and the original model with the AHRQ SES index variable added is 0.780.

Current submission

To further understand the relative importance of these risk-factors in the measure, we compared model performance with and without the addition of each social risk variable. Results show that the c-statistic is unchanged with the addition of any of these variables into the model: The c-statistic of the original model is 0.738; the c-statistic of the original model with the dual eligible variable added is 0.738; the original model with the AHRQ SES index variable added is 0.739; and the original model with the race (Black) variable added is 0.738.

Model	C-statistic
Baseline (no social risk factor)	0.738
Baseline plus Dual Eligible	0.738
Baseline plus low AHRQ SES	0.739
Baseline plus Race, (Black)	0.738

Table 12: C-statistics and predictive value for models with and without social risk factors

Analysis #5: Impact on hospital-level measure scores

Previous submission

We also examined the change in hospitals' RSMRs with the addition of any of these variables. The median absolute change in hospitals' RSMRs when adding a dual eligibility indicator is 0.010% (IQR: 0.00% – 0.03%, minimum 0.00% – maximum 0.48%) with a correlation coefficient between RSMRs for each hospital with and without dual eligibility added of 0.99781. The median absolute change in hospitals' RSMRs when adding an indicator for a low AHRQ SES Index score is 0.02% (IQR: 0.01% – 0.03%, minimum 0.00% – maximum 0.22%) with a correlation coefficient between RSRRs for each hospital with and without an indicator for a low AHRQ SES Index score added of 0.99961.

Current submission

For our updated analyses, we examined both correlations of measure scores, and calculated changes in measure scores.

- Dual eligibility: The median absolute change in hospitals' RSMRs when adding a dual eligibility indicator is 0.0% (IQR: 0.00% – 0.001 %, minimum 0.0% – maximum 7%) with a correlation coefficient between RSMRs for each hospital with and without dual eligibility added of 1.

- Low AHRQ SES: The median absolute change in hospitals' RSMRs when adding an indicator for a low AHRQ SES Index score is 0.027% (IQR: 0.012% – 0.066 %, minimum 0.00% – maximum 7%) with a correlation coefficient between RSRRs for each hospital with and without an indicator for a low AHRQ SES Index score added of 0.978.
- Race (Black): The median absolute change in hospitals' RSMRs when adding race (Black) is 0.0% (IQR: 0.00% – 0.001 %, minimum 0.0% – maximum 7%) with a correlation coefficient between RSMRs for each hospital with and without race (Black) added of 1.

Analysis #6: Relationship between measure scores and proportion of patients with social risk factors

To understand the impact of social risk factor adjustment on facilities that serve higher proportions of patients with social risk factors, we examined the relationship between measure scores and the proportion of patients with social risk factors for facilities in the highest quintile (for the facility- proportion of patients with social risk factors). Figures 6a, 6b, and 6c below show that relationship for each of the three social risk factors (dual eligible, low AHRQ SES, and race (Black)). For the dual eligible (Figure 6a) and Black variables (Figure 6c), there is no correlation across facilities, and specifically no positive correlation in the fifth quartile (in fact for dual eligibility there is a significant negative correlation indicating better measure scores for facilities treating a higher proportion of dual eligible patients). For the low AHRQ SES variable (Figure 6b) there is no significant association in the fifth quintile.

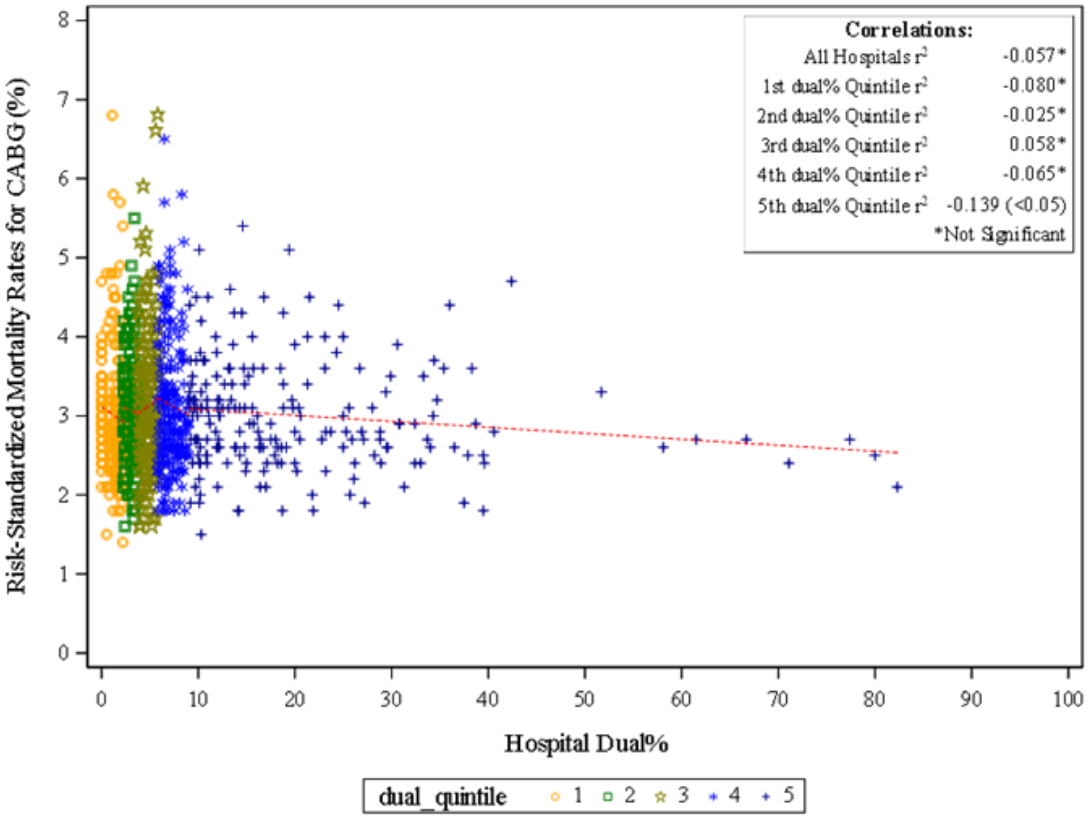


Figure 5: Relationship between facility proportion of patients with Dual Eligibility and measure scores

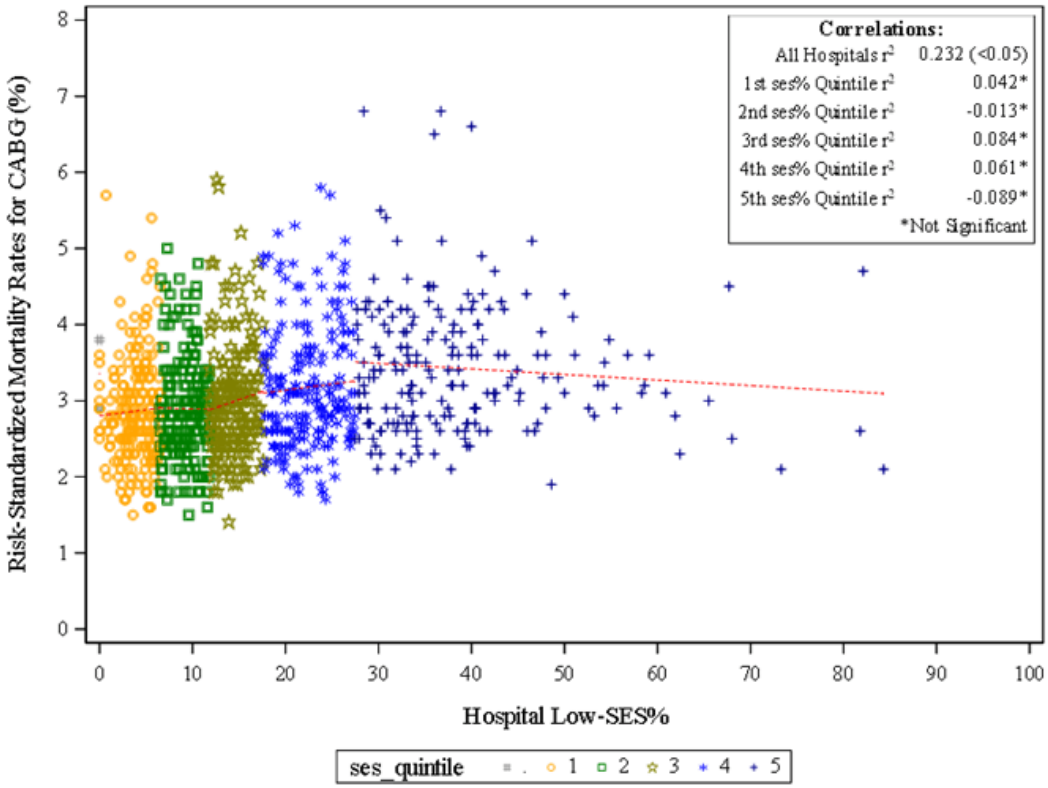


Figure 6: Relationship between facility proportion of patients with low AHRQ SES and measure scores

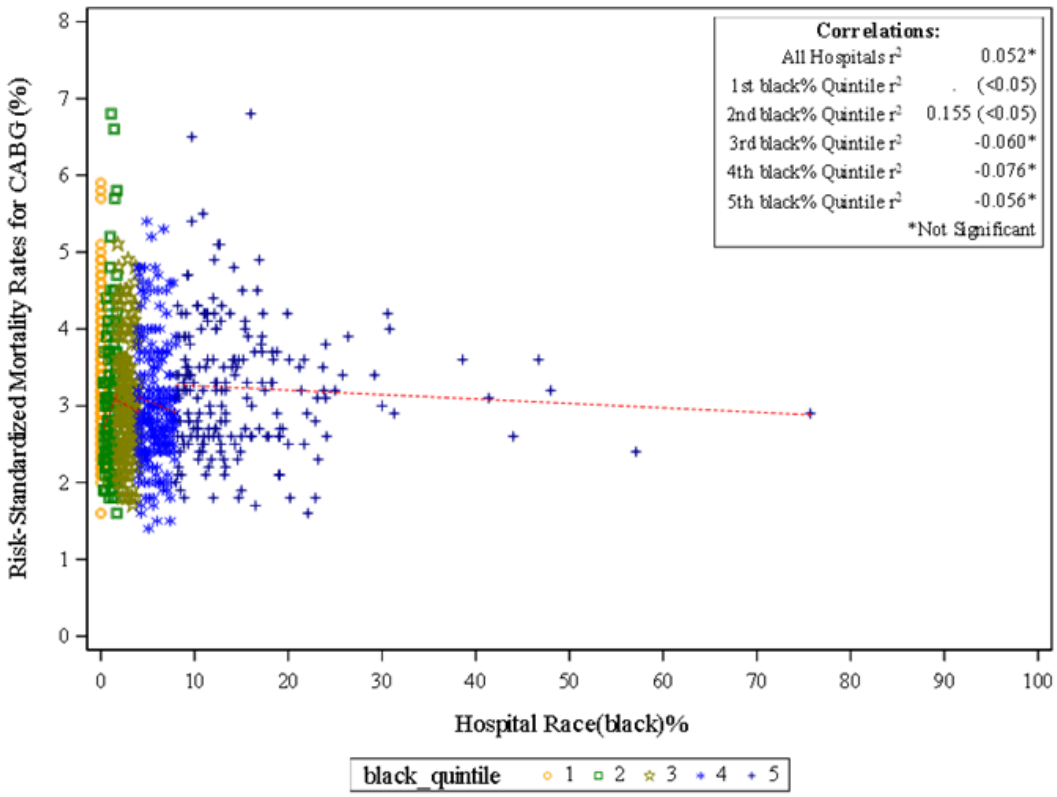


Figure 7: Relationship between facility proportion of Black patients and measure scores

Analysis #7: Assessing hospital- vs. patient-level effects of social risk factors

The results above show that of the three variables we tested, the low AHRQ SES variable is the only one to be both significantly associated with the outcome in a multivariate model, and for which the facility-proportion shows an association with measure scores.

One concern with including social risk factors in a model is that their effect may be at either the patient or the hospital level. For example, social risk factors may increase the risk of death because patients with social risk factors have an individual higher risk (patient-level effect) or because patients with social risk factors are more often admitted to hospitals with higher overall mortality rates (hospital-level effect). Identifying the relative contribution of the hospital level is important in considering whether a factor should be included in risk adjustment; if an effect is primarily a hospital-level effect, adjusting for it is equivalent to adjusting for differences in hospital quality. Thus, as an additional step, we assessed whether there was a "contextual effect" at the hospital level. To do this, we performed a decomposition analysis to assess the independent effects of the social risk variables at the patient level and the hospital level. If, for example, the elevated risk of death for patients with social risk factors were largely due to lower quality/higher mortality risk in hospitals with more patients with social risk factors, then a significant hospital-level effect would be expected with little-to-no patient-level effect. However, if the increased mortality risk were solely related to higher risk for patients with social risk factors regardless of hospital effect, then a significant patient-level effect would be expected and a significant hospital-level effect would not be expected.

To assess the relative contributions of the patient- vs hospital-level effects of this variable, we calculated the predicted probability of mortality for an "average" patient who went to hospitals with a high proportion of low AHRQ SES patients with the probability of mortality for an "average" patient who went to hospitals with a low proportion of low AHRQ SES patients (High proportion is defined as the 95th percentile of hospital proportion of low AHRQ SES patients; low proportion is 5th percentile). The difference in predicted probabilities at the 95th and 5th percentiles (P95-P5) estimates the hospital-level effect of the social risk factor. The difference in predicted probabilities when all patients have and do not have the social risk factor (delta) estimates the patient-level effect

of the SES risk factor on death. The hospital-level effect is greater than the patient-level effect when P95-P5 is greater than delta. We used P95 and P5 rather than the maximum (P100) and minimum (P0) to avoid outlier values.

Figure 8 shows that for the low AHRQ SES variable, there is a small patient-level effect (light blue bar) but a much larger hospital-level effect (dark blue bar), indicating that adjusting for this variable would reduce the hospital-level effect.

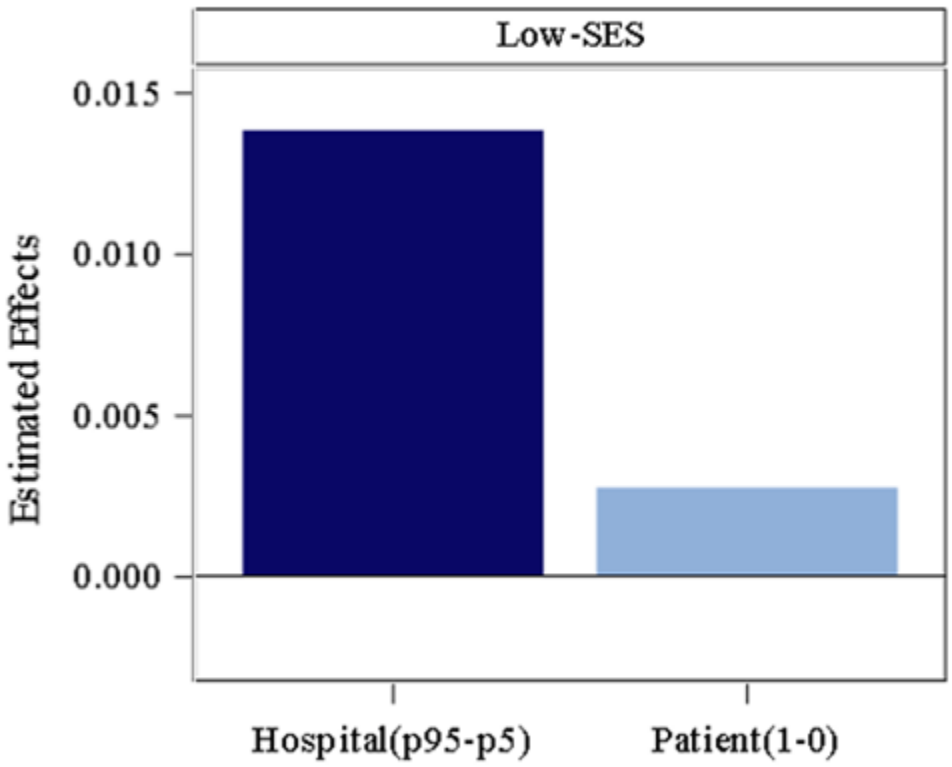


Figure 8: Hospital vs. patient-level effects of the low AHRQ SES Index variable for the 30-day CABG mortality outcome

Social risk factor testing: Conclusion

Previous submission

Overall, we find that the social risk variables that could be feasibly incorporated into this model do have a significant relationship with the outcome in multivariable modeling. However, the impact of any of these indicators is very small to negligible on model performance and hospital profiling. Given the controversial nature of incorporating such variables into a risk-model we do not support doing so in a case that is unlikely to affect hospital profiling.

Current submission

In summary, our results show that:

- Social risk factors are associated with higher observed outcomes for all three variables we tested (dual eligible, race (Black), low AHRQ SES) and higher odds ratios for the outcome in a bivariate model.

- In a multivariate model that includes all of the measures' risk variables, only the low AHRQ SES variable is statistically significantly associated with the outcome, suggesting that the clinical risk variables are attenuating the impact of the dual eligible and race (Black) variables.
- Measure scores calculated with and without the social risk factors are highly correlated, with very small absolute differences.
- None of the social risk variables show a significant relationship between the proportion of patients with that variable and measure scores across all facilities with the highest proportion of patients with social risk factors, including the low AHRQ SES variable.
- The impact of the low AHRQ SES variable is mainly at the hospital-level rather than the patient level.
- The risk model performs well (see calibration curves in the section below) separately for patients with each social risk factor.

In summary, the empiric results do not support including dual eligible, low AHRQ SES, or race (Black) in the risk model. Adjusting for low-AHRQ SES would mask a hospital-level effect, and while the dual eligible and race (Black) variables were not significantly associated with the outcome in a multivariable model, CMS feels it is not appropriate to add these variables to the risk model given the potential unintended consequences of signaling that differential care is acceptable. CMS believes that stratification of measure score results is the most acceptable approach for quality measures and has therefore contracted with CORE to develop two approaches at uncovering disparities in care. The *within hospital disparity method* compares outcomes for patients with social risk factors compared with patients without social risk factors receiving care within the same hospital; the risk model for this approach uses the same risk factors as the overall measure model, but also includes the social risk factor and the proportion of patients at the hospital with the social risk factor. The *across hospitals disparity method* calculates measure scores for patients with one social risk factor to allow for comparison across hospitals; the risk model accounts for differences in clinical case mix/medical conditions so that hospitals can be compared fairly. CMS has applied these methods to the readmission measures, and hospitals have recently received confidential reports with their results. CMS does not currently have plans to apply these methods to the mortality measures, however. We note that we are currently working on re-selecting risk variables for this and other CMS measures and this topic will be revisited by CORE, CMS, and CORE's risk-variable reselection TEP.

References:

Assessing Hospital Disparities for Dual Eligible Patients in 30-Day Condition-and Procedure-Specific Readmission Measures. (2021). Yale New Haven Health Services Corporation/Center for Outcomes Research & Evaluation.

Lloren A, Liu S, Herrin J, et al. Measuring hospital-specific disparities by dual eligibility and race to reduce health inequities. Health Services Research. 2019;54:243-254. doi:10.1111/1475-6773.13108

[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]

We computed three summary statistics for assessing model performance (Harrell and Shih, 2001) for the CABG mortality cohort:

Discrimination Statistics

(1) Area under the receiver operating characteristic (ROC) curve (the c-statistic) is the probability that predicting the outcome is better than chance, which is a measure of how accurately a statistical model is able to distinguish between a patient with and without an outcome)

(2) Predictive ability (discrimination in predictive ability measures the ability to distinguish high-risk subjects from low-risk subjects; therefore, we would hope to see a wide range between the lowest decile and highest decile.)

Calibration Statistics

(3) Over-fitting indices (over-fitting refers to the phenomenon in which a model accurately describes the relationship between predictive variables and outcome in the development dataset but fails to provide valid predictions in new patients)

We tested the performance of the model for **Dataset 1** described in section 2b.27.

CORE notes that after initial measure development we do not re-test our risk models for overfitting using a dataset that is external to the testing sample. In our risk models, coefficients are updated each time the measure is calculated. Therefore, random statistical fluctuations in model coefficients across repeated reporting cycles are part of the overall random error in the facility performance estimates. CORE believes that this approach is not a validity issue for this type of model, unlike the case of a static risk model.

Reference:

Harrell FE and Shih YC. Using full probability models to compute probabilities of actual interest to decision makers, Int. J. Technol. Assess. Health Care 17 (2001), pp. 17–26.

[Response Ends]

2b.27. Provide risk model discrimination statistics.

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

[Response Begins]

Previous submission

Results for the development cohort (Dataset 2)

2009 development cohort:

C-statistic = 0.75

Predictive ability (lowest decile %, highest decile %): (0.7, 11.1)

2008 validation cohort:

C-statistic = 0.74

Predictive ability (lowest decile %, highest decile %): (0.6, 11.8)

2010 validation cohort:

C-statistic = 0.75

Predictive ability (lowest decile %, highest decile %): (0.5, 10.6)

Results for the 2017 reporting cohort (Dataset 1)

C statistic = 0.7789;

Predictive ability (lowest decile %, highest decile %) = (0.4, 14.0)

Current submission

Results for the 2022 EM Dataset

C-statistic: 0.74

Predictive ability (lowest decile %, highest decile %) = (0.8, 10.2)

[Response Ends]

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

[Response Begins]

Results for the development cohort (Dataset 2)

2009 development cohort: Calibration (over-fitting statistics): (0, 1)

2008 validation cohort: Calibration (over-fitting statistics): (0.01, 0.99)

2010 validation cohort: Calibration (over-fitting statistics): (-0.10, 0.97)

[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]

The risk decile plot is a graphical depiction of the deciles calculated to measure predictive ability. Below, we present the risk decile plot showing the distributions for Medicare FFS data.

Previous submission

Dataset 1 (July 2013-June 2019)

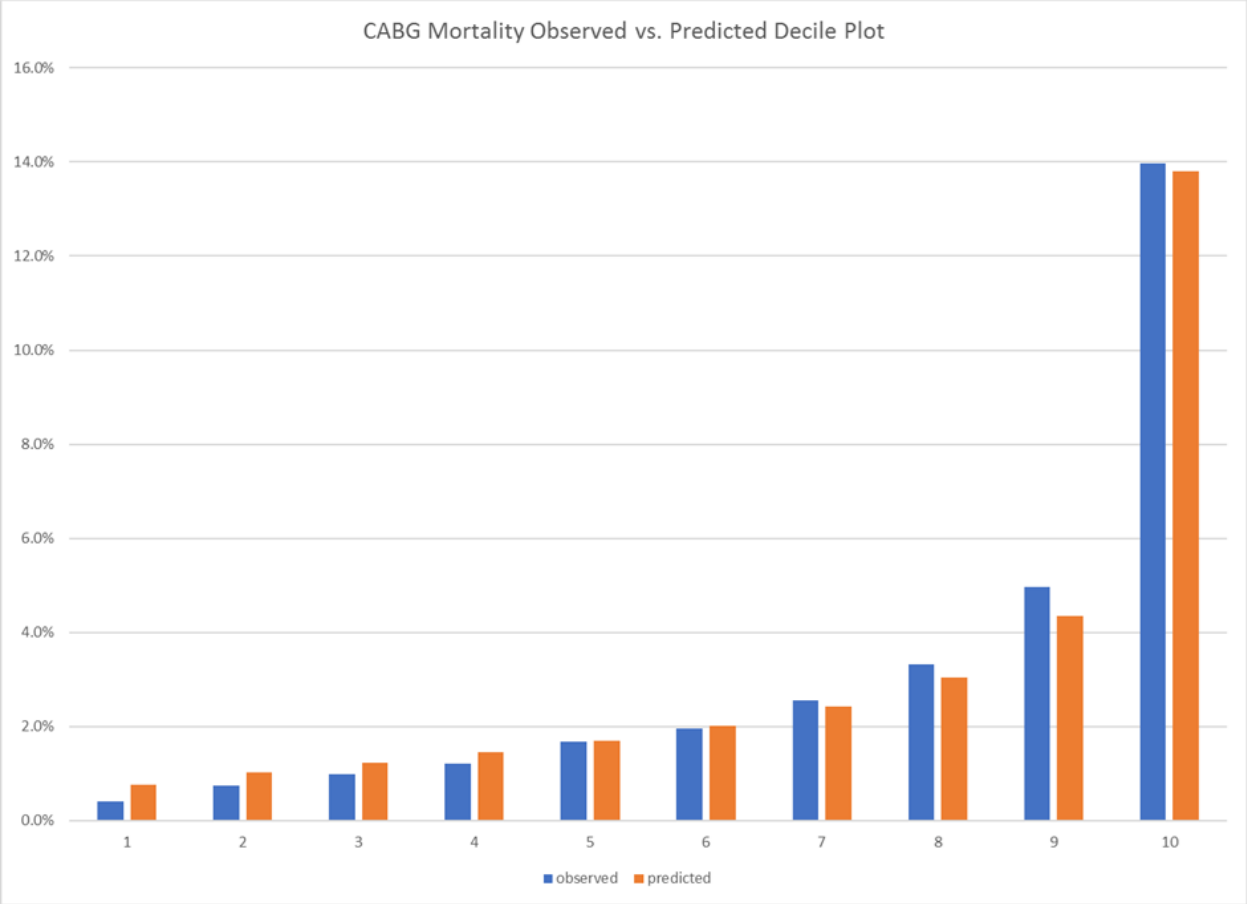
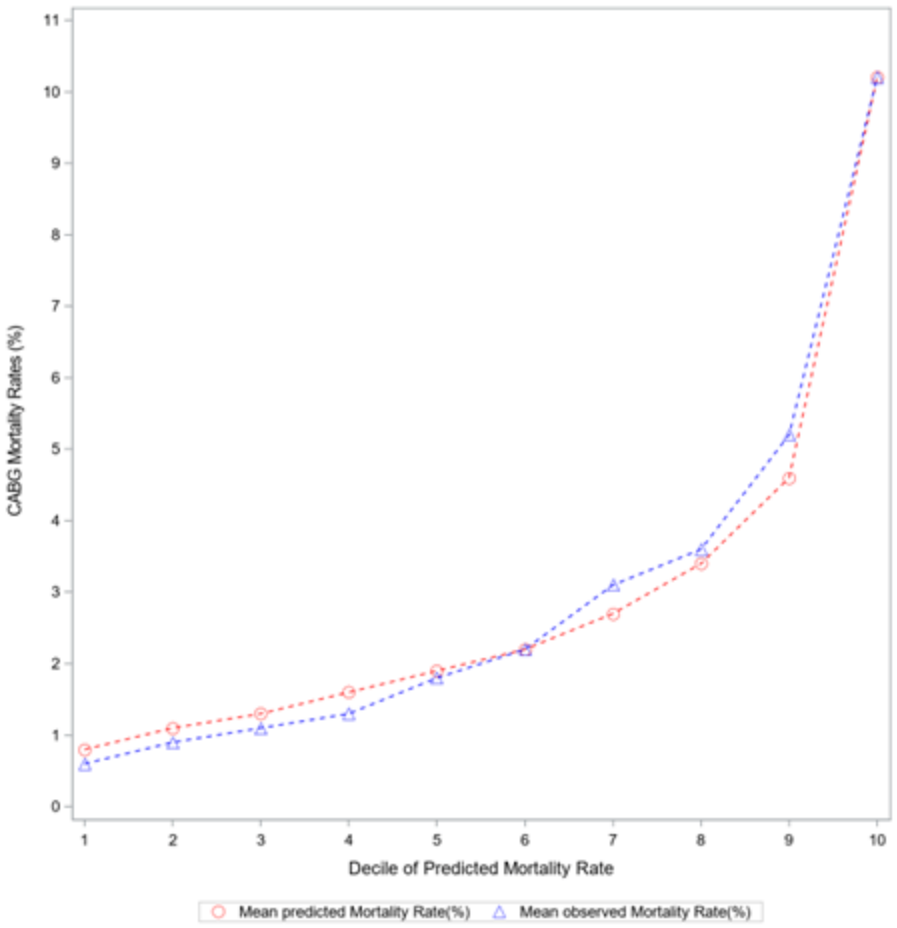


Figure 9. Risk-decile plot for all patients in cohort

Current submission

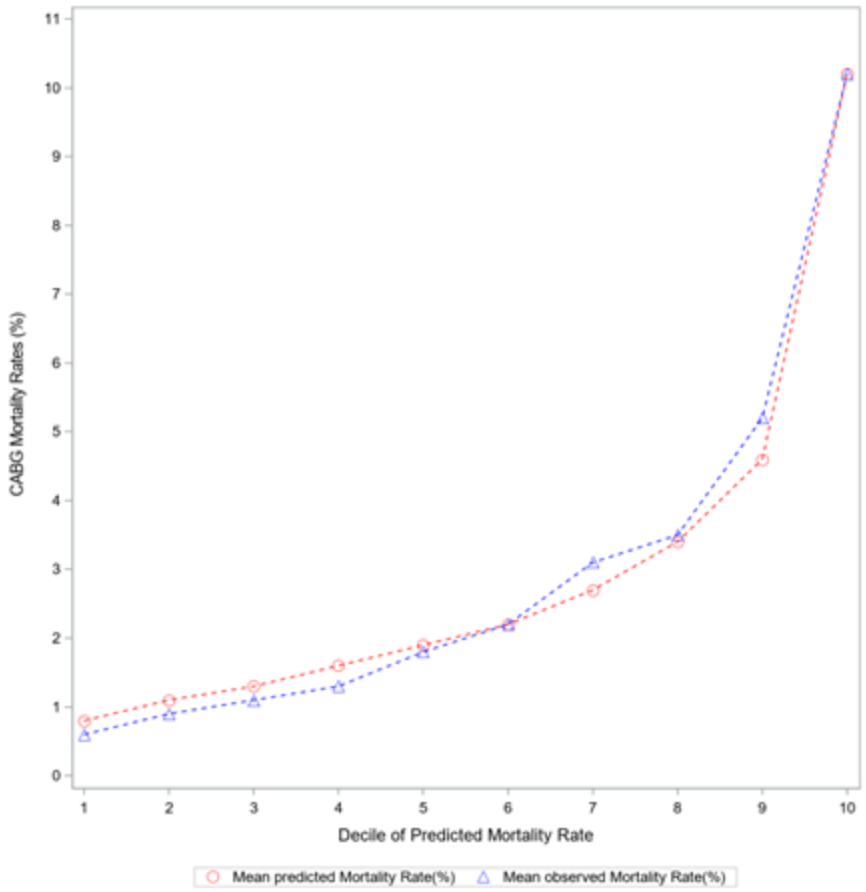
2022 EM Dataset (July 2016-June 2019)

Below we provide risk-decile plots for all patients (Figure 10), as well as for patients with dual eligibility, low AHRQ SES, and race (Black) (Figures 11, 12, 13).



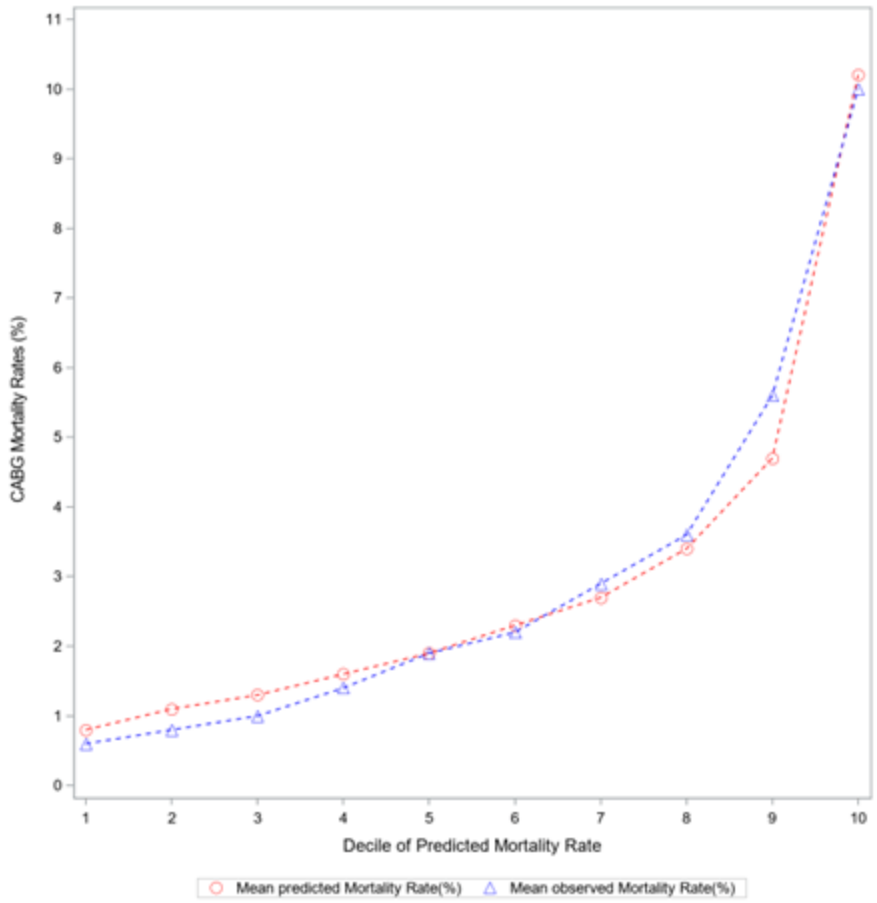
Note:
Among all patients.

Figure 10: Risk-decile plot, all patients



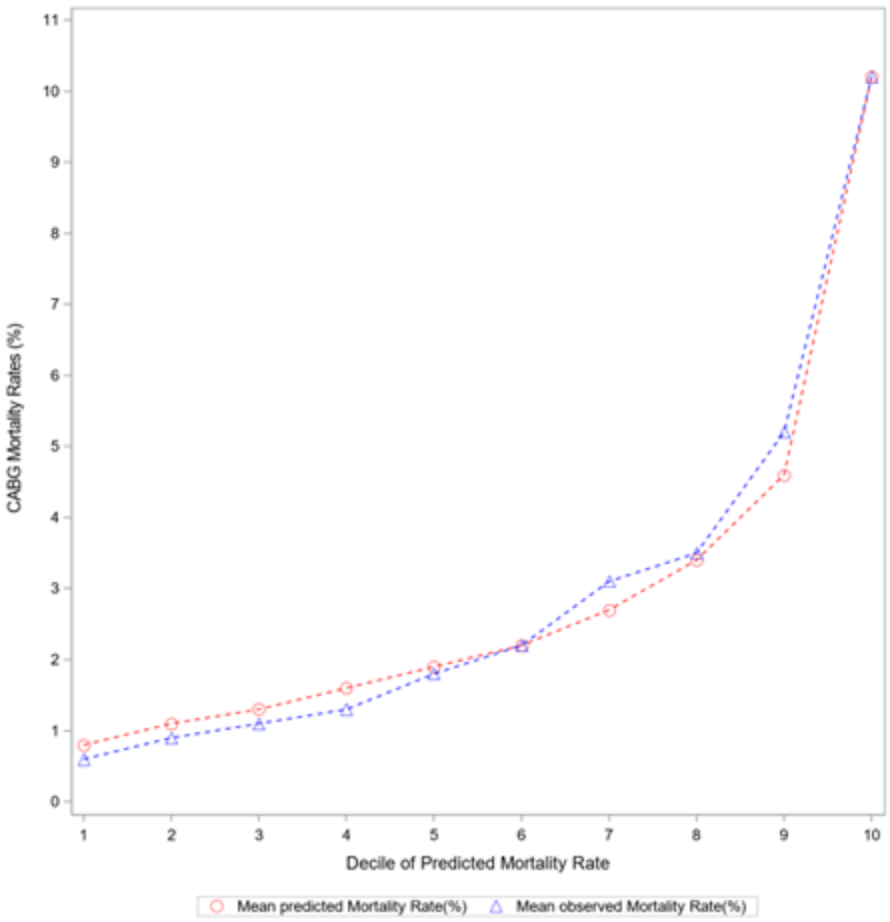
Note:
Among patients with Dual Eligible for Medicare and Medicaid.

Figure 11: Risk-decile plot, patients with Dual Eligibility



Note:
Among patients with AHRQ Low SES (9-digit ZIP).

Figure 12: Risk-decile plot, patients with low AHRQ SES Index



Note:
Among patients with Race (Black).

Figure 13: Calibration curve, Black patients

[Response Ends]

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

[Response Begins]

This measure is not stratified.

[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]

Discrimination Statistics (Dataset 1 and Dataset 2)

The C-statistics ranged from 0.75 to 0.78 across datasets (Dataset 1 and Dataset 2) and indicates good model discrimination. The model indicated a wide range between the lowest decile and highest decile, indicating the ability to distinguish high-risk subjects from low-risk subjects.

Calibration Statistics (Dataset 2)

Over-fitting (Calibration γ0, γ1)

If the γ0 in the validation samples are substantially far from zero and the γ1 is substantially far from 1, there is potential evidence of over-fitting. The calibration value of close to zero at one end and close to 1 on the other end indicates good calibration of the model (Dataset 2).

Risk Decile Plots (Dataset 1)

Higher deciles of the predicted outcomes are associated with higher observed outcomes, which show a good calibration of the model. This plot indicates excellent discrimination of the model and good predictive ability.

Overall Interpretation (Dataset 1 and Dataset 2)

Interpreted together, our diagnostic results demonstrate the risk-adjustment model adequately controls for differences in patient characteristics (case mix).

Current submission:

Our results show that the risk adjustment model continues to adequately control for differences in patient characteristics, with good calibration, excellent discrimination, and good predictive ability.

CORE notes that after initial measure development we do not re-test our risk models for overfitting using a dataset that is external to the testing sample. In our risk models, coefficients are updated each time the measure is calculated. Therefore, random statistical fluctuations in model coefficients across repeated reporting cycles are part of the overall random error in the facility performance estimates. CORE believes that this approach is not a validity issue for this type of model, unlike the case of a static risk model.

[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]

No additional information.

[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]

Coded by someone other than person obtaining original information (e.g., DRG, ICD-10 codes on claims)

[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]

ALL data elements are in defined fields in electronic claims

[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]

N/A

[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]

There are no fees associated with the use of this measure.

[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]

[Response Ends]

4a.02. Check all planned uses.

[Response Begins]

Public reporting

Payment Program

[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. This measure is currently publicly reported.

[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. This measure is currently publicly reported.

[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]

The exact number of measured entities (acute care hospitals) varies with each new measurement period. In 2019, 1,179 hospitals were included in measure calculation, and results were publicly reported for 1,015 hospitals (Wallace et al., 2019). These include admissions to non-federal short-term acute care hospitals (including Indian Health Service hospitals) and critical access hospitals with at least 25 cases performed between July 2015 and June 2018.
Each hospital receives their measure results in April of each calendar year through CMS's QualityNet website. The results are then publicly reported on CMS's Hospital Compare website in July of each calendar year. Since the measure is risk-standardized using data from all hospitals, hospitals cannot independently calculate their score. However, CMS provides each hospital with several resources that aid in the interpretation of their results (described in detail below). These include Hospital-Specific Reports (HSRs) with details about every patient from their facility that was included in the measure calculation (for example, dates of admission and discharge, discharge diagnoses, outcome [died or not], transfer status, and facility transferred from). These reports facilitate quality improvement (QI) activities such as review of individual deaths and patterns of deaths; make visible to hospitals post-discharge outcomes that they may otherwise be unaware of; and allow hospitals to look for patterns that may inform QI work (e.g. among patient transferred in from particular facilities).
The HSRs also provide hospitals with more detailed benchmarks with which to gauge their performance relative to peer hospitals and interpret their results, including comorbidity frequencies for their patients relative to other hospitals in their state and the country.
Additionally, the code used to process the claims data and calculate measure results is written in SAS (Cary, NC) and is provided each year to hospitals upon request to make the measure methodology completely transparent.
Reference:
1. Wallace L, Grady J, Djordjevic D, et al. 2019 Procedure-Specific Measures Updates and Specifications Report Hospital-Level 30-Day Risk-Standardized Mortality Measure.
https://qualitynet.org/inpatient/measures/mortality/methodology

[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]

In April of each year, hospitals have access to the following list of updated resources related to the measure which is provided directly or posted publicly for hospitals to use:
1. HSRs: available for hospitals to download from QualityNet in April of each calendar year; includes information on the index admissions included in the measure calculation for each facility, detailed measure results, and state and national results.
2. HSR User Guide: available with the HSR and posted on QualityNet; provides instructions for interpreting the results and descriptions of each data field in the HSR.
3. Mock HSR: posted on QualityNet; provides real national results and simulated state and hospital results for stakeholders who do not receive an HSR.

4. Inpatient Quality Reporting (IQR) Preview Reports and Preview Report Help Guide: available for hospitals to download from QualityNet in April of each calendar year; includes measure results that will be publicly reported on Hospital Compare.
5. Measure Updates and Specification Reports: posted in April of each calendar year on QualityNet with detailed measure specifications, descriptions of changes made to the measure specifications with rationale and impact analysis when appropriate, updated risk variable frequencies and coefficients for the national cohort, and updated national results for the new measurement period.
6. Frequently asked Questions (FAQs): includes general and measure-specific questions and responses, as well as infographics that explain complex components of the measure's methodology, and are posted in April of each calendar year on QualityNet.
7. The SAS code used to calculate the measure results with documentation describing what data files are used and how the SAS code works. The SAS code and documentation are updated each year and are released upon request beginning in July of each year.
8. Measure Fact Sheets provide a brief overview of measures, measure updates, and are posted in April of each calendar year on QualityNet.
In July of each year, the publicly-reported measure results are posted on Hospital Compare, a tool to find hospitals and compare their quality of care that CMS created in collaboration with organizations representing consumers, hospitals, doctors, employers, accrediting organizations, and other federal agencies. Measure results are updated in July of each calendar year.

[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]

Questions and Answers (Q&A)
The measured entities (acute care hospitals) and other stakeholders or interested parties submit questions or comments about the measure through an email inbox (CMSreadmissionmeasures@yale.edu). Experts on measure specifications, calculation, or implementation, prepare responses to those inquiries and reply directly to the sender. We consider issues raised through the Q&A process about measure specifications or measure calculation in measure reevaluation.

[Response Ends]

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

[Response Begins]

Summary of Questions or Comments from Hospitals submitted through the Q & A process:
For the CABG mortality measure, we have received the following inquiries from hospitals since the completion of measure maintenance in March 2018:
1. Requests for detailed measure specifications including ICD-10 codes and Condition Categories used to define the measure cohort or in the risk-adjustment model;
2. Requests for the SAS code used to calculate measure results;
3. Queries about how cohorts and outcomes are defined;
4. Questions about the cohort inclusion and exclusion criteria;
5. Requests for hospital-specific measure information, such as data included in the HSRs.

[Response Ends]

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

[Response Begins]

Summary of Question and Comments from Other Stakeholders:
For the CABG mortality measure, feedback received from other stakeholders since the submission of the last annual form in December 2016 has included the following:
1. Requests for detailed measure specifications including ICD-9 and ICD-10 codes used to define the measure cohort or in the risk-adjustment model;
2. Requests for the SAS code used to calculate measure results; and
3. Requests for clarification of how inclusion and exclusion criteria are applied, such as if hospice patients are included in the measure cohort;

Summary of Relevant Publications from the Literature Review:
Since December 2015, we have reviewed 8 articles related to mortality following isolated CABG surgery. One article examined CABG mortality disparities using a similarly measure, but no articles employed the measure score results.

References:
1. RH Mehta, DM Shahian, S Sheng, SM O'Brien, FH Edwards, JP Jacobs and ED Peterson. Association of Hospital and Physician Characteristics and Care Processes with Racial Disparities in Procedural Outcomes Among Contemporary Patients Undergoing Coronary Artery Bypass Grafting Surgery. Circulation. 2016;133:124-130, http://dx.doi.org/10.1161/CIRCULATIONAHA.115.015957

[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]

Each year issues raised through the Q&A or in the literature related to this measure are considered by measure and clinical experts. Any issues that warrant additional analytic work due to potential changes in the measure specifications are addressed as a part of annual measure reevaluation. If small changes are indicated after additional analytic work is complete, those changes are usually incorporated into the measure in the next measurement period. If the changes are substantial CMS may propose the changes through rulemaking and adopt the changes only after CMS received public comment on the changes and finalizes those changes in the Inpatient Prospective Payment System (IPPS) or other rule. There were no questions or issues raised by stakeholders requiring additional analysis or changes to the measure since maintenance of endorsement in 2018. There have been no changes made to the measure since endorsement maintenance in 2018.

[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]

The median hospital 30-day, all-cause, RSMR for the CAGB mortality measure for the 3-year period between July 1, 2013 and June 30, 2016 was 3.1%. The median RSMR decreased by 0.1 absolute percentage points from July 2013-June 2014 (median RSRR: 3.1%) to July 2015-June 2016 (median: RSRR: 3.0%).

[Response Ends]

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

[Response Begins]

N/A

[Response Ends]

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

[Response Begins]

N/A

[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]

N/A

[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]

We did not include in our list of related measures any non-outcome (e.g., process) measures with the same target population as our measure. Our measure cohort was heavily vetted by clinical experts, a technical expert panel, and a public comment period. In addition, the related claims-based CABG readmission measure, which utilizes the same definition of isolated CABG as the mortality measure, was validated using STS clinical registry data. Because this is an outcome measure, clinical coherence of the cohort takes precedence over alignment with related non-outcome measures. Furthermore, non-outcome measures are limited due to broader patient exclusions. This is because they typically only include a specific subset of patients who are eligible for that measure (for example, patients who receive a specific medication or undergo a specific procedure).

[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]

The NQF-endorsed STS measure that has the same target population and similar measure focus as the proposed CABG mortality measure is the Risk-adjusted operative mortality for CABG (NQF #0119). The measure steward for the registry-based mortality measure for CABG is STS. In developing the measure, we sought to harmonize with the STS measure to the greatest extent feasible given competing measure design objectives and differences in the data source. The potential sources of discrepancy are target patient population, age, isolated CABG, period of observation, and included hospitals. The STS measure also assesses both deaths occurring during CABG hospitalization (in-hospital death, even if after 30 days) and deaths occurring within 30 days of procedure date. As indicated above, the proposed measure uses a standard follow-up period of 30 days of procedure date in order to measure each patient consistently. The proposed claims-based measure has been tested and is appropriate for use in all-payer data for patients 18 years and over. Finally, the STS cardiac surgery registry currently enrolls most, but not all, patients receiving CABG surgeries in the U.S. The proposed CABG mortality measure will capture all qualifying Medicare FFS patients undergoing CABG regardless of whether their hospital or surgeon participates in the STS registry.

[Response Ends]

Appendix

Supplemental materials may be provided in an appendix.:

Available in attached file

Attachment: 2558_2020_PSM_v1.0.pdf

Contact Information

Measure Steward (Intellectual Property Owner): Centers for Medicare & Medicaid Services

Measure Steward Point of Contact: Dollar-Maples, Helen, helen.dollar-maples@cms.hhs.gov

Measure Developer if different from Measure Steward: Yale New Haven Health Services Corporation – Center for Outcomes Research and Evaluation (CORE)

Measure Developer Point(s) of Contact: Peter, Doris, doris.peter@yale.edu

Dollar-Maples, Helen, helen.dollar-maples@cms.hhs.gov

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]

Available in attached file

[Response Ends]

Attachment: 2558_2020_PSM_v1.0.pdf

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

Describe the members' role in measure development.

[Response Begins]

Technical Expert Panel Members:
Joseph V. Agostini, MD, Aetna
Tanya Alteras, MPP, National Partnership for Women and Families
Mary Barton, MD, MPP, National Committee for Quality Assurance (NCQA)
Carol Beehler, RN, NEA-BC, Pricewaterhouse Coopers
Todd Michael Dewey, MD, Southwest Cardiothoracic Surgeons
Lee Fleisher, MD (Served from March 30, 2012 to May 25, 2012), American Society of Anesthesiologists, University of Pennsylvania School of Medicine
Paul Kurlansky, MD, Florida Heart Research Institute, Inc
Frederic Masoudi, MD, MSPN, University of Colorado-Denver, Senior Medical Office of National CV Data Registries
Christine McCarty, MD, Cardiovascular Surgical Institute
Joseph Parker, PhD, State of California: Office of Statewide Health Planning and Development,
Kenneth Sands, MD, MPH, Beth Israel Deaconess Medical Center
Ed Savage, MD, Cleveland Clinical Florida
Stephen Schmaltz, PhD, The Joint Commission
Richard Shemin, MD, UCLA Medical Center
Alan Speir, MD, Inova Fairfax Hospital

Working Group Panel Members:
Arnar Geirsson, MD, Yale School of Medicine
David Shahian, MD, STS Workforce on National Databases, Harvard Medical School, Massachusetts General Hospital

[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]

Annual

[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]

N/A

[Response Ends]

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

[Response Begins]

N/A

[Response Ends]

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

[Response Begins]

N/A

[Response Ends]