MEASURE WORKSHEET

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

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

Brief Measure Information

NQF #: 2515

De.2. **Measure Title:** Hospital 30-day, all-cause, unplanned, risk-standardized readmission rate (RSRR) following coronary artery bypass graft (CABG) surgery

Co.1.1. **Measure Steward:** Centers for Medicare & Medicaid Services

De.3. **Brief Description of Measure:** The measure estimates a hospital-level risk-standardized readmission rate (RSRR), defined as unplanned readmission for any cause within 30 days from the date of discharge of the index CABG procedure, for patients 18 years and older discharged from the hospital after undergoing a qualifying isolated CABG procedure. The measure was developed using Medicare Fee-for-Service (FFS) patients 65 years and older and was tested in all-payer patients 18 years and older. An index admission is the hospitalization for a qualifying isolated CABG procedure considered for the readmission outcome.

1b.1. **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 readmission 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 readmission is a priority area for outcome measure development, as it is an outcome that is likely attributable to care processes and is an important outcome for patients. Measuring and reporting readmission rates will inform healthcare providers and facilities about opportunities to improve care, strengthen incentives for quality improvement, and ultimately improve the quality of care received by Medicare patients. The measure will also provide patients with information that could guide their choices, as well as increase transparency for consumers.

S.4. **Numerator Statement:** The outcome for this measure is 30-day all-cause readmission. We define all-cause readmission as an unplanned inpatient admission for any cause within 30 days after the date of discharge from the index admission for patients 18 years and older who were discharged from the hospital after undergoing isolated CABG surgery. If a patient has one or more unplanned admissions (for any reason) within 30 days after discharge from the index admission, only one is counted as a readmission.

S.7. **Denominator Statement:** This claims-based measure can be used in either of two patient cohorts: (1) patients aged 65 years or older or (2) patients aged 18 years or older. We have tested the measure in both age groups.

The cohort includes admissions for patients a) who receive a qualifying isolated CABG procedure and b) with a complete claims history for the 12 months prior to admission. For simplicity of implementation and as testing demonstrated, closely correlated patient-level and hospital-level results using models with or without age interaction terms, the only recommended modification to the measure for application to all-payer data sets is replacement of the "Age-65" variable with a fully continuous age variable.

S.10. **Denominator Exclusions:** In order to create a clinically coherent population for risk adjustment and in accordance with existing NQF-approved CABG measures and clinical expert opinion, the measure is intended to capture isolated CABG patients (i.e., patients undergoing CABG procedures without concomitant valve or other major cardiac or vascular procedures).

For all cohorts, hospitalizations are excluded if they meet any of the following criteria, for admissions:

1. Without at least 30 days post-discharge enrollment in FFS Medicare
2. Discharged against medical advice (AMA)
3. Admissions for subsequent qualifying CABG procedures during the measurement period

De.1. **Measure Type:** Outcome

S.23. **Data Source:** Claims (Only)
S.26. **Level of Analysis:** Facility

IF Endorsement Maintenance – Original Endorsement Date: Dec 23, 2014 **Most Recent Endorsement Date:** Dec 23, 2014

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

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

De.4. **IF PAIRED/GROUPED, what is the reason this measure must be reported with other measures to appropriately interpret results?** This measure is not formally paired with another measure, however this measure is harmonized with a measure of hospital-level, all-cause, 30-day, risk-standardized mortality following a qualifying isolated CABG procedure.

Preliminary Analysis

Criteria 1: Importance to Measure and Report

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

1a. Evidence

1a. Evidence. The evidence requirements for a health outcomes measure include providing rationale that supports the relationship of the health outcome to processes or structures of care. The guidance for evaluating the clinical evidence asks if the relationship between the measured health outcome and at least one clinical action is identified and supported by the stated rationale.

Evidence Summary

- The developer states a number of recent studies have demonstrated that improvements in care at the time of patient discharge can reduce 30-day readmission rates. The developer noted a variety of research studies that revealed readmission rates are influenced by the quality of care provided within the health system and, specifically, that interventions such as improved discharge planning, reconciling patient medications, and improving communications with outpatient providers can reduce readmission rates.
- The developer noted this readmission measure was developed to identify institutions, whose performance is better or worse than expected based on patient case-mix.

Changes to evidence from last review

☒ **The developer attests that there have been no changes in the evidence since the measure was last evaluated.**
☐ **The developer provided updated evidence for this measure:**

Question for the Committee:

○ *Is there at least one intervention that a provider can undertake to achieve a change in the measure results?*

Guidance from the Evidence Algorithm
Box 1: The measure assesses performance on a health outcome → Box 2: There is a relationship between the heath outcome and healthcare action → Pass

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

1b. Gap in Care/Opportunity for Improvement and 1b. Disparities
Maintenance measures – increased emphasis on gap and variation

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

- In 2007, CABG was ranked as one having the highest potentially preventable readmission rate within 15 days following discharge (13.5%) by the Medicare Payment Advisory Committee (MedPAC) in a report to Congress.
- In applying this measure to Medicare claims data from 2009-2011, the developer stated the range in hospital-level RSRRs is 13.3% to 21.3%, indicating performance variation among measured entities. The median RSRR is 16.8% (25th and 75th percentiles are 15.6% and 17.9%, respectively). The distribution of RSRRs across hospitals is shown below:

	RSRR (%)
Maximum	23.1
90%	19.2
75%	17.9
Median	16.8
25%	15.6
10%	14.6
Minimum	12.0

- **Disparities:** The developer conducted analyses to explore disparities in hospitals' performance on the CABG readmission measure by race and socioeconomic status (SES).
- In regard to race, the hospitals with fewer African-American patients performed slightly better than hospitals with a higher number of African-American patients, but the two groups show a similar range of performance. See table provided by the developer below.

Note: In table below, %AA = proportion of African American patients

Decile	#Hospitals	%AA(min)	%AA(max)	RSRR(median)	RSRR(min)	RSRR(max)
.	1,197	0	100	16.85%	12.53%	22.36%
3	358	0	0.86	16.76%	12.66%	21.88%
4	121	0.86	1.67	16.71%	12.85%	21.09%
5	120	1.68	2.53	16.74%	13.30%	21.11%
6	119	2.53	3.72	17.21%	13.80%	21.53%
7	119	3.73	5.73	16.71%	14.11%	21.93%
8	121	5.75	8.74	16.93%	13.58%	22.07%
9	120	8.75	13.91	17.21%	14.04%	21.77%
10	119	13.96	100	17.34%	12.53%	22.36%

- The developers used Medicaid data to indicate SES performance and found that hospitals with the most Medicaid beneficiaries perform slightly worse than hospitals with the fewest Medicaid beneficiaries, but the two groups show a similar range of performance. See table provided by the developer below.

Decile	# Hospitals	%Medicaid (min)	%Medicaid (max)	RSRR (median)	RSRR (min)	RSRR (max)
.	1,197	0	100	16.85%	12.53%	22.36%
1	119	0	3.26	16.85%	13.30%	21.88%
2	119	3.27	5.15	16.40%	12.66%	21.10%
3	116	5.17	6.65	16.53%	13.48%	20.20%
4	125	6.67	7.86	16.85%	13.62%	21.93%
5	119	7.87	9.23	16.58%	13.34%	22.07%
6	121	9.26	11.1	16.75%	12.53%	21.68%

7	118	11.18	13.61	16.97%	13.81%	20.79%
8	121	13.64	18.55	17.37%	13.35%	22.36%
9	120	18.56	29.36	16.94%	12.76%	21.77%
10	119	29.41	100	17.44%	14.11%	21.11%

Questions for the Committee:

- *Is there a gap in care that warrants a national performance measure?*

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

Committee pre-evaluation comments

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

1a. Evidence to Support Measure Focus

If measuring a structure, process, or intermediate outcome: How does the evidence relate to the specific structure, process, or intermediate outcome being measured? Does it apply directly or is it tangential? How does the structure, process, or intermediate outcome relate to desired outcomes?

If measuring a health outcome or PRO: is the relationship between the measured outcome/PRO and at least one healthcare action (structure, process, intervention, or service) identified AND supported by the stated rationale?

Comments:

** Pass

** Evidence indicates that the outcome of readmission rates are influenced by the quality of care and clinical actions within health systems. It recognizes a broad view of quality that goes beyond individual process of care measures and involves communication between providers, how complications are handled, coordination of transitions, etc.

** Yes, although the extent to which the measure and data used are able to capture structures/processes/interventions etc. employed by providers is unclear. For example, a dichotomous variable derived from administrative data cannot directly measure structures such as "coordinated transitions to the outpatient environment". While there is undoubtedly a relationship between readmission and transitional care coordination, the measure does not account for differences in access to points of formal and informal care available to patients across communities, post-discharge.

** Pass

** Agreed: evidence exists to support utility of measure. Lived experience since readmissions focus began would support that there are multiple interventions that can improve readmission rates

** Readmission rates are a well-established quality measure in the hospital quality measurement and improvement programs. There remains some debate as to whether they truly measure hospital quality (with a few recent publications demonstrating little relationship between processes of care and readmission rates for medical conditions). However, incenting hospitals to focus on readmissions does seem to have encouraged them to begin to think outside their walls.

** The evidence is related to the health outcome measured and health care actions were identified.

** No changes

** Data does demonstrate the relationship between intervention and outcome.

** The measure is designed to identify hospitals that are performing worse or better than expected based on their case mix.

** Recent articles suggest good transitions of care can affect readmission for these patients

** Yes

** Readmission rates can be influenced by the quality of care provided...interventions could include transitions of care, discharge planning, medication reconciliation, communication improvement that could help to reduce readmissions.

1b. Performance Gap

Was performance data on the measure provided? How does it demonstrate a gap in care (variability or overall less than optimal performance) to warrant a national performance measure? Disparities: Was data on the measure by population subgroups provided? How does it demonstrate disparities in the care?

Comments:

** Performance gap: Yes; Opportunity for improvement: yes

** CABG is considered as having significant preventable readmission rates and performance variation among entities that have been measured. There appears to be opportunity for improvement.

** Not enough information was provided to answer meaningfully.
1) It's unclear whether this endorsement is for the HRRP CABG measure, all-payer 18+ from California, or both. If the latter, it seems as though two individual evaluations would be more appropriate--for example, the national Medicare FFS age 65+ model was presumably originally tuned with great attention to identifying clinically-relevant risk adjusters (expert panel with candidate measures retained with bootstrapped stepwise validity tests, etc.). Do those results hold up in an all-payer setting of 18+ Californians?
2) As currently-specified the models produce meaningful variation in risk-standardized performance assessments, however failing to account for social risk would lead many to believe the models are mis-specified. More justification is needed for excluding the few SDS factors they tested than citing similar ranges among deciles. "

** High

** There is a performance gap: between lowest performing and highest performing hospitals nearly two fold difference. On historical claims data, gap is somewhat less but still significant. Re: disparities, developers ranked hospitals by AA% (as proxy for race) and Medicaid% (as proxy for SES). Both factors seem to slightly negatively impact performance but absolute differences are small.

** Readmission rates remain higher than optimal, although there is likely not a ""zero floor"" for this clinical outcome. Not all readmissions are preventable.

There are disparities for this readmission measure, particularly for dually-enrolled beneficiaries. Racial disparities are minimal."

** There is variation in performance sufficient to warrant a national measure. Disparities data was analyzed and some disparate performance was observed.

** Developer suggested performance gap between race and Medicaid eligibility

** Good analysis of SDS including dual eligible status.

** The developer cited studies relating certain interventions at the time of discharge with improved readmissions rates. The developer did note a range in performance amongst measured hospitals.

** MedPac report to Congress in 2007 regarding a high potential for preventable readmission exists.
Risk adjustments indicate a range of 23.1 - 12% rate. Indicates race and payer slight disparities"

** Yes

** Range in hospital level RSRR is from 13.3% to 21.3%
Also noted differences based on race and SES"

1c. Composite Performance Measure – Quality Construct

Are the following stated and logical: overall quality construct, component performance measures, and their relationships; rationale and distinctive and additive value; and aggregation and weighting rules?

N/A

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications
Maintenance measures – no change in emphasis – specifications should be evaluated the same as with new measures

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

Data source(s):

- CMS Administrative Claims

Specifications:

- The numerator for this measure includes all-cause readmissions that are unplanned for any cause within 30 days after the date of discharge from the index admission for patients 18 years and older who were discharged from the hospital after undergoing isolated CABG surgery. If a patient has one or more unplanned admissions (for any reason) within 30 days after discharge from the index admission, only one is counted as a readmission.
- The denominator for this measure can be either of two patient cohorts: (1) patients aged 65 years or older or (2) patients aged 18 years or older. The developer tested the measure in both age groups. The cohort includes admissions for patients a) who receive a qualifying isolated CABG procedure and b) with a complete claims history for the 12 months prior to admission.
- Hospitalizations are excluded if they meet any of the following criteria, for admissions:
  1. Without at least 30 days post-discharge enrollment in FFS Medicare
  2. Discharged against medical advice (AMA)
  3. Admissions for subsequent qualifying CABG procedures during the measurement period
- The measure is specified for a facility level of analysis and the hospital setting
- The measure is specified with a statistical risk model with 26 risk adjustment factors.
  - The measure employs a hierarchical logistic regression model to create a hospital-level 30-day RSRR.
  - The developer notes that this approach simultaneously models two levels (patient and hospital) to account for the variance in patient outcomes within and between hospitals.

Questions for the Committee:

- *Are all the data elements clearly defined? Are all appropriate codes included?*
- *Is the logic or calculation algorithm clear?*
- *Is it likely this measure can be consistently implemented?*

2a2. Reliability Testing, Testing attachment
Maintenance measures – less emphasis if no new testing data provided

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers.

For maintenance measures, summarize the reliability testing from the prior review:

- The reliability testing submitted by the measure developer from the prior review is fairly consistent with the information presented to the Standing Committee for maintenance review. There are no material updates to the reliability testing.

SUMMARY OF TESTING

Reliability testing level ☐ **Measure score** ☐ **Data element** ☒ **Both**

Reliability testing performed with the data source and level of analysis indicated for this measure ☐ **Yes** ☐ **No**

Method(s) of reliability testing

- Data Source: Medicare inpatient and outpatient claims across 2008-2010 was used to test the reliability of the measure.
- There were 175,891 admissions in the three year sample.

Data element testing

- The developers state that they tested the face validity of the measure's critical data elements using the CMS audit process to ensure accuracy of claims coding as these data elements are consequential for payment. NQF guidelines require a systematic assessment of face validity. NQF requires a systematic and transparent process to evaluate the face validity by experts who are not involved in measure development.
- The developers also compared variable frequencies and odds ratios from logistic regression models across the three years of data.

Measure score testing

- The developers take a "test-retest" approach to measuring reliability. The developers randomly spilt the dataset into two equal subsets and calculated the RSRR for each sample. The developers use a metric of agreement known as an intra-class correlation coefficient (ICC) to measure agreement between the two samples.

Results of reliability testing

Data element testing

- The developers do not provide results of a systematic assessment of face validity.
- The developers note that there is little change in risk factor frequencies across the three year period.

Measure score testing

- The inter-class correlation coefficient between the two RSRRs for each hospital was 0.331
- The developer notes that this result can be interpreted as "fair" agreement.

Questions for the Committee:

- *Is the test sample adequate to generalize for widespread implementation?*
- *Are the methods and results of data element reliability testing robust?*
- *Is a ICC of 0.331 sufficient to demonstrate measure score reliability?*

Guidance from the Reliability Algorithm

1. Specifications are precise (YES) → 2. Empirical Reliability testing conducted (YES) → 3. Testing was computed at the performance score level (YES) → 5. The testing method appropriate (YES) → 6b. Testing results demonstrate moderate confidence in measure score reliability → Rating: Moderate

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

2b. Validity
Maintenance measures – less emphasis if no new testing data provided

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

Specifications consistent with evidence in 1a. ☒ **Yes** ☐ **Somewhat** ☐ **No**

Question for the Committee:

- *Are the specifications consistent with the evidence?*

2b2. Validity testing

2b2. Validity Testing should demonstrate the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality.

For maintenance measures, summarize the validity testing from the prior review:

- The validity testing submitted by the measure developer from the prior review is fairly consistent with the information presented to the Standing Committee for maintenance review. There are no material updates to the validity testing.

SUMMARY OF TESTING

Validity testing level ☒ **Measure score** ☐ **Data element testing against a gold standard** ☐ **Both**

Method of validity testing of the measure score:

☒ **Face validity only**

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

Validity testing method:

- The developer describes several validity tests. First, the developer asserts the validity of claims-based measures noting that prior measures for alternate conditions have been endorsed and used for public reporting. Prior measures have been tested against their authoritative source to demonstrate that the underlying data elements are valid. However, NQF requirement require validity testing be conducted with the measure as specified.
- The developer notes that the measure is valid since it was developed based on measure development guidelines. While following measure development guidelines is highly encouraged, NQF requires testing on either data elements or the measure score.
- The developer explains that the measure was assessed by external groups providing results of as systematic assessment of face validity. The developers surveyed their technical expert panel. A systematic assessment of face validity generally requires an assessment of experts not involved in the development of the measure.
- Finally, the developer evaluates the validity of the measure cohort and risk adjustment model with registry data validation.

Validity testing results:

- The systematic assessment of face validity demonstrated that 71% of the measure developers technical advisory panel agreed that the measure will provide an accurate reflection of quality.
- The registry validation of the patient cohort demonstrated an overall agreement rate of 95.6% of matched patients between the claims cohort and the registry cohort.
    - The developer notes that any inconsistencies between the two cohorts can be due to coding errors in the claims data, abstraction errors in the registry data, or may be due to inconsistencies in the probabilistic matching process used to create a matched set of patients for the validation.
- To demonstrate the validity of the risk adjustment model, the developer compared the distribution of hospital RSRRs with the claims-based and registry-based measures. The developer found that overall 63 out of 829 (7.6%) of the hospitals had greater than 1% absolute difference in RSRR between the claims-based vs. registry-based measure. In particular, 8 hospitals changed performance categories. Note, these results are only generalizable to STS hospitals and the STS registry does not capture all patients in all hospitals.

Questions for the Committee:

- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient validity so that conclusions about quality can be made?*
- *Do you agree that the score from this measure as specified is an indicator of quality?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

- The measure includes three exclusions:

- Hospital stays in which patients leave hospital against medical advice (AMA)
- Hospital stays for patients without at least 30 days post-discharge information
- Subsequent hospital stays for patients with additional CABG procedure admissions within 30 days

- The measure exclusions represent a small number of patients in the sample used by the developers

Questions for the Committee:

- *Are any patients or patient groups inappropriately excluded from the measure?*

2b4. Risk adjustment: **Risk-adjustment method** ☐ **None** ☒ **Statistical model** ☐ **Stratification**

Conceptual rationale for SDS factors included ? ☒ **Yes** ☐ **No**

SDS factors included in risk model? ☒ **Yes (gender)** ☐ **No**

Risk adjustment summary

- The measure includes a statistical risk model with 26 risk factors
- The measure employs a hierarchical logistic regional model (HGLM) to create the hospital-level 30-day RSRR.
- The risk adjustment model includes demographic factors (age, gender), and markers of comorbidity and disease severity. The table below summarizes the risk factors used in the model:

Demographics
Mean age minus 65 (SD)
Male (%)
Comorbidities
History of Coronary Artery Bypass Graft (CABG) or valve surgery (ICD-9 diagnosis codes: V42.2, V43.3, V45.81, 414.02, 414.03, 414.04, 414.05, 414.06, 414.07, 996.02, 996.03; ICD-9 procedure code: 39.61)
Cardiogenic shock (ICD-9 diagnosis code 785.51)
Chronic Obstructive Pulmonary Disease (COPD) (CC 108)
Cancer; metastatic cancer and acute leukemia (CC 7-12)
Diabetes mellitus (DM) or DM complications (CC 15-19, 119-120)
Protein-calorie malnutrition (CC 21)
Disorders of fluid/electrolyte/acid-base (CC 22-23)
Other endocrine/metabolic/nutritional disorders (CC 24)
Severe hematological disorders (CC 44)
Dementia or other specified brain disorders (CC 49-50)
Major psychiatric disorders (CC 54-56)
Hemiplegia, paraplegia, paralysis, functional disability (CC 67-69, 100-102, 177-178)
Polyneuropathy (CC 71)
Congestive heart failure (CC 80)
Specified arrhythmias and other heart rhythm disorders (CC 92-93)
Stroke (CC 95-96)
Cerebrovascular disease (CC 97-99, 103)
Vascular or circulatory disease (CC 104-106)
Fibrosis of lung or other chronic lung disorders (CC 109)
Pneumonia (CC 111-113)
Other lung disorders (CC 115)
Dialysis status (CC 130)

- The developer tested three SDS and race variables in their analysis: dual eligible status, African American race,

AHRQ SES index.

- These variables were tested based on four potential pathways that were considered:
  - Relationship of socioeconomic status factor to health at admission
  - Use of low-quality hospital
  - Differential care within a hospital
  - Influence of SES on readmission risk outside of hospital quality and health status
- When the SDS and race variables were tested in a multivariate model, the effect size of each of the variables was modest. The c-statistic was unchanged, and the model with the SDS factors had little to no effect on hospital performance.
- The developers also undertook a decomposition analysis. They found that patient-level race and low AHRQ SES index effects were not appreciably different from zero. However, hospital-level race and low AHRQ SES effects were significant. The table is provided here:

CABG Readmission Decomposition Analysis

Parameter	Estimate (Standard Error)	P-value
Dual Eligible – Patient-Level	0.1705 (0.0269)	<.0001
Dual Eligible – Hospital-Level	0.3400 (0.1467)	0.0205
African American – Patient-Level	0.0067 (0.0347)	0.8472
African American – Hospital-Level	0.5452 (0.1403)	0.0001
AHRQ SES Index – Patient-Level	0.0357 (0.0202)	0.0777
AHRQ SES Index – Hospital-Level	0.2185 (0.0512)	<.0001

- Given these findings and the complex pathways, the developers did not incorporate the SDS and race variables into the measure.

- The metric for determining risk model discrimination is the c-statistic. The c-statistic is a measure of goodness of fit in a logistic regression model. The c-statistic gives the probability that a randomly selected patient who experienced a readmission had a higher risk score than a patient who had not experienced the readmission. The range for c-statistics is 0.5 to 1. The c-statistic for this risk model was 0.62.
- The calibration statistics demonstrated a value of close to zero at one end and close to one on the other end. This was consistent with the 2009 development cohort, 2008 validation cohort and the 2010 validation cohort.

Questions for the Committee:

- *Is an appropriate risk-adjustment strategy included in the measure?*
- *Are the candidate and final variables included in the risk adjustment model adequately described for the measure to be implemented?*
- *Do you agree with the developer's decision, based on their analysis, to not include SDS factors in their risk-adjustment model?*

2b5. Meaningful difference *(can statistically significant and clinically/practically meaningful differences in performance measure scores can be identified)*:

- The developer used the January 2009-September 2011 cohort and found that the risk-adjusted range of performance for hospitals was 12.0% to 23.1%, with the 25th-75th percentile ranging from 15.6-17.9%.

Question for the Committee:

- *Does this measure identify meaningful differences about quality?*

2b6. Comparability of data sources/methods:

N/A

2b7. Missing Data

N/A

Guidance from the Validity Algorithm

Precise specifications (Box 1) → Empirical Validity Testing on the measure as specified (Box 6) → moderate certainty that the measure score is reliable

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

Committee pre-evaluation comments

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

2a1. & 2b1. Specifications

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

Comments:

** Clearly defined: yes; Logic clear: yes; Can be consistently implemented: yes

** I think there is likelihood that this measure can be consistently implemented and relies on administrative claims data that is routinely collected.

** The developers are very thorough on the specifications of the Medicare cohort. Unsure on the California cohort.

** Specs reliability: acceptable

** Similar measure specs to other RSRR measures: claims data, all cause readmits with few exclusions, hierarchical regression model. Reliability specs and testing were previously presented.

** Elements of risk adjustment model are relatively simplistic - yes/no on a number of fairly broad clinical comorbidities. Unclear if better data on the comorbidities would improve measure performance, however. Specifications are clear and this is a fairly easily-implemented measure.

** None

** 2 cohorts (65+ yoa and 18+ yoa)
Developers did not offer why they had 2 grps. "

** ICD9 codes used in data review. Need to see data using identified ICD10 codes to determine ease of implementation.

** None

** 26 risk factors, clear rationale for numerator and denominator

** Appropriate

** Inter-class correlation coefficient was 0.331 (fair)
Seems lower than you would want for a measure

2b.1 Validity

In what ways, if any, are the specifications inconsistent with the evidence? If a PRO-PM: In what ways, if any, are the specifications inconsistent with what the target population values and finds meaningful?

Comments:

**Consistent with evidence: yes

** If consumers are the target population, some have argued that a stratified approach is most meaningful (Zaslavsky, Jha).

If the target population is policy makers/CMS, the approach is consistent if the model is properly specified by accounting for factors not mediated by hospital quality, subject to existing reimbursement parameters (e.g. pay for an inhaler, but not a home assessment to prevent triggers).

** Specs validity: consistent

** Specs are consistent with evidence

** 1) risk-adjustment components, as above, could likely benefit from being more comprehensive and/or nuanced to clinical severity. The issue of disparities, particularly for duals, remains - is this due to inadequate capture of medical risk, or due to independent effect of social factors, social support, etc.?
2) use of a shrinkage model in the setting of a procedure for which there is a proven volume-outcome relationship always warrants discussion, and this is no exception.

** No issues other than defining ICD10 codes.

** None

**Asserts validity of claims based measures since measure was tested using measure specifications.

** Yes

** Results only generalizable to STS hospitals and STS registry does not capture all patients in all hospitals which may limit validity of measure

2a.2 Reliability

Was reliability tested with an adequate scope (number of entities and patients) to generalize for widespread implementation and with an appropriate method? Describe how the results either do or do not demonstrate sufficient reliability. If a PRO-PM: Was testing conducted at both the data element and score levels? If a composite: Was testing conducted at the score level?

Comments:

** Reliability testing: yes; Reliability rating: moderate

** The measure was developed using Medicare FFS patients 65 years and older and tested with all-payer patients 18 and older.

** Some would argue an ICC of .3 indicates poor agreement.

** Acceptable- moderate

** Test-retest method results characterized as "fair" by developers, but ICC of 0.331 is categorized as "poor" by other sources.

** Testing demonstrated "fair" reliability. I'd like to understand more about the potential causes of low reliability before commenting on whether this could be improved.

** Yes

** Claims data used was old, from 2008-2010. Would like to see if the same performance exists with more current data. There have been significant changes both in the care provides as well as patient selection over the years so not confident that the data is still representative.

** The developer used a large sample of ~176k admissions across three years.

** Tested critical elements using the CMS audit process. Used test, retest approach splitting the group prior to calculating RSRR. .Noted little change over 3 years and "fair" agreement

** Yes - although this will be challenged

** Inter-class correlation coefficient was 0.331 (fair)
Seems lower than you would want for a measure"

2b2. Validity

Was validity tested with an adequate scope (number of entities and patients) to generalize for widespread implementation and with an appropriate method? Describe how the results either do or do not demonstrate sufficient validity so that conclusions about quality can be made? Why do you agree (or not agree) that the score from this measure as specified is an indicator of quality? If a PRO-PM: Was testing conducted at both the data element and score levels?

Comment:

** Test sample adequate: yes; Results valid: yes; Score indicates quality: yes

** Testing appeared to be adequate.

** Again depends on larger questions around specification.

** Acceptable- moderate

** Face validity analysis conducted via survey of developers' TEP which may not be adequate per NQF guidelines.

** Validity testing as reported compares whether or not the current measure matches up with a clinical measure in terms of who was readmitted, but does not address the issues of causality, whether risk is adequately taken into account, or shrinkage.

** Yes

** No issues noted.

** There were no material updates to the validity results

** Face validity is 71% - of technical advisory panel
Over 95% overall agreement in matched patient
Risk model has limitations to STS hospitals

** Exclusions only represent a small number of patients

2b3.-2b7. Threats to Validity

2b3. Exclusions: Are the exclusions consistent with the evidence? Are any patients or patient groups inappropriately

excluded from the measure? Are the exclusions/exceptions of sufficient frequency and variation across providers to be needed (and outweigh the data collection burden)?

2b4. Risk Adjustment: If outcome (intermediate, health, or PRO-based) or resource use performance measure: Is there a conceptual relationship between potential SDS variables and the measure focus? How well do SDS variables that were available and analyzed align with the conceptual description provided? Are all of the risk-adjustment variables present at the start of care (if not, do you agree with the rationale provided)?. Was the risk adjustment (case-mix adjustment) appropriately developed and tested? Do analyses indicate acceptable results? Is an appropriate risk-adjustment strategy included in the measure?

2b5. Meaningful Differences: How do analyses indicate this measure identifies meaningful differences about quality?

2b6. Comparability of performance scores: If multiple sets of specifications: Do analyses indicate they produce comparable results? If risk-adjustment approach includes SDS factors: Did the developer compare performance scores with and without SDS factors in the risk-adjustment approach? Did the results support the risk-adjustment approach?

2b7. Missing data/no response: Does missing data constitute a threat to the validity of this measure?

Comments:

** Patients appropriately excluded: yes; Risk-adjustment strategy appropriate: yes; Variables adequately described: Yes; Agree with decision not to include SDS factor: yes; Measure score is reliable: yes

** It appears that hospitals with higher Medicaid, African-American and dual eligible populations have higher readmission rates. However, they have similar ranges of performance indicating that these hospitals can also perform well on this measure.

it was hard to untangle patient-level variables, community level variables and hospital-level variables.

Given the findings and the complex pathways that could explain any relationship between SES or race with readmissions, neither were incorporated into the measure.

** I see question 2b4. above as most critical, but there is no text box to provide feedback. Is the role of SDS out of scope for this evaluation? If not, I feel the measure deserves a more thorough evaluation of SDS than is provided in the information here. For example, the coefficient for hospital-level race (0.5452) would yield an odds ratio of 1.725 which would be the strongest predictor in the model (compared to the clinical covariates on pg. 48). What impact did the tested SDS variables have on hospital assessments and were there differences for groups of hospitals (i.e. safety net, etc.)? They also assume linearity in the area deprivation index--other studies have shown effects are most pronounced in the tails (Kind, Jencks, et al). What about interacting community and patient-level SDS factors to evaluate different dimensions of poverty for different populations? Additionally ZIP codes can be quite large and subject to ecological fallacy. Would it be possible to test effects of more granular measures of community deprivation?

** Acceptable modeling

** 2b3. Reasonable exclusions, relatively few in sample
2b4. developers tested 3 SDS variables: dual eligibility, AA race SES index. Inclusion in model did not significantly change c-statistic and did not impact hospital performance. Reluctantly agree with developer's decision to remove race/SDS from measure based on data presented. Still wonder if there is something else we are not capturing well.
2b5. There are significant differences in performance between best/worst scoring hospitals.

** Missing data is not a major issue.

** N/A

** The lack of a more recent data set is concerning regarding any changes over time.

** None

** Opted not to include SDS and race although there was some modest change in results

** Noted as rare

2d. Composite Performance Measure

Do analyses demonstrate the component measures fit the quality construct and add value? Do analyses demonstrate the aggregation and weighting rules fit the quality construct and rationale?

N/A

Criterion 3. Feasibility
Maintenance measures – no change in emphasis – implementation issues may be more prominent

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

[feasibility]

- This measure is calculated using administrative claims data from defined data fields in electronic claims. Thus, the measure's required data elements are routinely collected as part of the facilities billing process.
- There are no fees, licensing, or other requirements to use any aspect of the measure as specified.

Questions for the Committee:

- *Is the data collection strategy ready to be put into operational use?*

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

Committee pre-evaluation comments

Criteria 3: Feasibility

3. Feasibility

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

Comments:

** Feasible: Yes

** High

** As uses administrative claims data, no concerns re: feasibility

** Highly feasible measure

** Highly feasible, issues raised not applicable

** Able to capture data with appropriate ICD10 codes.

** None

** Administrative claims data easily accessible

** Administrative data

Criterion 4: Usability and Use
Maintenance measures – increased emphasis – much greater focus on measure use and usefulness, including both impact /improvement and unintended consequences

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

Current uses of the measure [from OPUS]
Publicly reported? ☒ **Yes** ☐ **No**

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

Accountability program details

- The measure is currently used in CMS' Hospital Inpatient Quality Reporting (IQR) Program. Based on the number of participating hospitals, the risk-standardized readmission rate (RSRR) was reported for 4,663 hospitals across the United States for 2015 public reporting. The final index cohort included 925,315 admissions.
- The measure has also been used in CMS' Hospital Readmission Reduction (HRRP) Program. The number of accountable entities participating in the HRRP program varies by reporting year.

Improvement results Developers found that the mean RSRR decreased from 15.0% between July 2012 and June 2013 to 13.9% between July 2014 and June 2015. The median hospital RSRR in the combined three-year dataset was 14.4%. These reductions indicate progress in 30-day RSRR for CABG.

Unexpected findings (positive or negative) during implementation N/A

Potential harms N/A

Vetting of the measure N/A

Feedback:

- This measure was originally endorsed in December 2014 and has not since undergone maintenance evaluation.

***Questions for the Committee**:*
- *How can the performance results be used to further the goal of high-quality, efficient healthcare?*
- *Do the benefits of the measure outweigh any potential unintended consequences?*

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

Committee pre-evaluation comments
Criteria 4: Usability and Use

4. Usability and Use

How is the measure being publicly reported? For maintenance measures – which accountability applications is the measure being used for? How can the performance results be used to further the goal of high-quality, efficient healthcare? Describe any actual unintended consequences and note how you think the benefits of the measure outweigh them.

Comments:

** Currently in use: yes; Unintended consequences: Hospitals may reduce CABG readmission rate while increasing post discharge mortality rate. Need to monitor and report both rates in tandem; Validation of readmission rate should include mortality as a measure variable.

** Developer indicates measure is not eligible for endorsement because the community of entities being measured has expressed concern about the unintended negative consequences of the measure. They point out that there is a growing body of literature that academic medical centers and safety net facilities are incurring penalties under the HRRP where measure is used disproportionate to other hospitals.

** Moderate

** Already in use and being publicly reported. Developers report on impact on improvement in performance.

** Highly useable measure. How it is used is more complicated. The measure was intended to identify outliers with specificity, rather than to grade hospitals across a range of performance - but it is used the former way in Hospital Compare and the latter in the HRRP.

** This measure is being used in the IQR and HRRP programs.

** Appreciate that SDS is more than dual eligible status recognizing that inner city hospitals are at a disadvantage when comparing readmission rates across all hospitals.

** The measure is currently in use by the CMS' IQR program.

** Measure currently used in IQR and HRRP programs - improvements realized.

** This is the issue, is it sensitive enough to discriminate small differences?

** Yes

Criterion 5: Related and Competing Measures

Related or competing measures
N/A

Harmonization
N/A

Endorsement + Designation

The "Endorsement +" designation identifies measures that exceed NQF's endorsement criteria in several key areas. After a Committee recommends a measure for endorsement, it will then consider whether the measure also meets the "Endorsement +" criteria.

This measure is a candidate for the "Endorsement +" designation IF the Committee determines that it: meets evidence for measure focus without an exception; is reliable, as demonstrated by score-level testing; is valid, as demonstrated by score-level testing (not via face validity only); and has been vetted by those being measured or other users.

Eligible for Endorsement + designation: ☐ **Yes** ☐☒ **No**

RATIONALE IF NOT ELIGIBLE:

The community of entities being measured has expressed concern about the unintended negative consequences of this measure. The measure is being scored in the HRRP program with an observed to expected ratio rather than an interquartile range. The entities being measures have also expressed concerns that there is a growing body of literature noting that academic medical centers and safety net facilities are incurring penalties under the HRRP at rate disproportionate to other hospitals.

Pre-meeting public and member comments

-

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** Hospital 30-day, all-cause, unplanned, risk-standardized readmission rate (RSRR) following coronary artery bypass graft (CABG) surgery

Measure Title: Click here to enter measure title

IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here: Click here to enter composite measure #/ title

Date of Submission: 1/11/2017

Instructions

- *Complete 1a.1 and 1a.12 for all measures.*
- *Complete **EITHER 1a.2, 1a.3 or 1a.4** as applicable for the type of measure and evidence.*
- *For composite performance measures:*
  - *A separate evidence form is required for each component measure unless several components were studied together.*
  - *If a component measure is submitted as an individual performance measure, attach the evidence form to the individual measure submission.*
- All information needed to demonstrate meeting the evidence subcriterion (1a) must be in this form. An appendix of *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.

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

1a. Evidence to Support the Measure Focus

The measure focus is evidence-based, demonstrated as follows:

- Health outcome: [3] a rationale supports the relationship of the health outcome to processes or structures of care. Applies to patient-reported

outcomes (PRO), including health-related quality of life/functional status, symptom/symptom burden, experience with care, health-related behavior.

- Intermediate clinical outcome: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured intermediate clinical outcome leads to a desired health outcome.
- Process: [5] a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured process leads to a desired health outcome.
- Structure: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured structure leads to a desired health outcome.
- Efficiency: [6] evidence not required for the resource use component.

Notes

3. Generally, rare event outcomes do not provide adequate information for improvement or discrimination; however, serious reportable events that are compared to zero are appropriate outcomes for public reporting and quality improvement.

4. The preferred systems for grading the evidence are the U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.

5. Clinical care processes typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with patient input) → provide intervention → evaluate impact on health status. If the measure focus is one step in such a multistep process, the step with the strongest evidence for the link to the desired outcome should be selected as the focus of measurement. Note: A measure focused only on collecting PROM data is not a PRO-PM.

6. Measures of efficiency combine the concepts of resource use and quality (see NQF's Measurement Framework: Evaluating Efficiency Across Episodes of Care; AQA Principles of Efficiency Measures).

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

Outcome

☒ Health outcome: 30-day, all-cause, unplanned, risk-standardized readmission rate (RSRR) following coronary artery bypass graft (CABG) surgery

☐ Patient-reported outcome (PRO): Click here to name the PRO

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

☐ Intermediate clinical outcome (*e.g., lab value*): Click here to name the intermediate outcome

☐ Process: Click here to name what is being measured

☐ Appropriate use measure: Click here to name what is being measured

☐ Structure: Click here to name the structure

☐ Composite: Click here to name what is being measured

1a.12 LOGIC MODEL Diagram or briefly describe the steps between the healthcare structures and processes (e.g., interventions, or services) and the patient's health outcome(s). The relationships in the diagram should be easily understood by general, non-technical audiences. Indicate the structure, process or outcome being measured.

The goal of this measure is to improve patient outcomes by providing patients, physicians, and hospitals with information about hospital-level, risk-standardized readmission rates following coronary artery bypass graft (CABG) surgery. 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 readmission measure was developed to identify institutions, whose performance is better or worse than would be expected based on their patient case-mix, and therefore promote hospital quality improvement and better inform consumers about care quality.

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

1a.2 FOR OUTCOME MEASURES including PATIENT REPORTED OUTCOMES- State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process (e.g., intervention, or service).

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. Furthermore, research on a variety of conditions and procedures has shown that readmission rates are influenced by the quality of care provided within the health system and, specifically, that interventions such as improved discharge planning, reconciling patient medications, and improving communications with outpatient providers can reduce readmission rates. A number of recent studies have demonstrated that improvements in care at the time of patient discharge can reduce 30-day readmission rates.[1-16]

References:

[1]Naylor M, Brooten D, Jones R, Lavizzo-Mourey R, Mezey M, Pauly M. Comprehensive discharge planning for the hospitalized elderly. A randomized clinical trial. *Ann Intern Med.* Jun 15 1994;120(12):999-1006.

[2]Naylor MD, Brooten D, Campbell R, et al. Comprehensive discharge planning and home follow-up of hospitalized elders: a randomized clinical trial. *Jama.* Feb 17 1999;281(7):613-620.

[3]Krumholz HM, Amatruda J, Smith GL, et al. Randomized trial of an education and support intervention to prevent readmission of patients with heart failure. *J Am Coll Cardiol.* Jan 2 2002;39(1):83-89.

[4]van Walraven C, Seth R, Austin PC, Laupacis A. Effect of discharge summary availability during post-discharge visits on hospital readmission. *J Gen Intern Med.* Mar 2002;17(3):186-192.

[5]Conley RR, Kelly DL, Love RC, McMahon RP. Rehospitalization risk with second-generation and depot antipsychotics. *Ann Clin Psychiatry.* Mar 2003;15(1):23-31.

[6]Coleman EA, Smith JD, Frank JC, Min S-J, Parry C, Kramer AM. Preparing patients and caregivers to participate in care delivered across settings: the Care Transitions Intervention. *J Am Geriatr Soc.* Nov 2004;52(11):1817-1825.

[7]Phillips CO, Wright SM, Kern DE, Singa RM, Shepperd S, Rubin HR. Comprehensive discharge planning with postdischarge support for older patients with congestive heart failure: a meta-analysis. *JAMA.* Mar 17 2004;291(11):1358-1367.

[8]Jovicic A, Holroyd-Leduc JM, Straus SE. Effects of self-management intervention on health outcomes of patients with heart failure: a systematic review of randomized controlled trials. *BMC Cardiovasc Disord.* 2006;6:43.

[9]Garasen H, Windspoll R, Johnsen R. Intermediate care at a community hospital as an alternative to prolonged general hospital care for elderly patients: a randomised controlled trial. *BMC Public Health.* 2007;7:68.

[10]Mistiaen P, Francke AL, Poot E. Interventions aimed at reducing problems in adult patients discharged from hospital to home: a systematic meta-review. *BMC Health Serv Res.* 2007;7:47.

[11]Courtney M, Edwards H, Chang A, Parker A, Finlayson K, Hamilton K. Fewer emergency readmissions and better quality of life for older adults at risk of hospital readmission: a randomized controlled trial to determine the effectiveness of a 24-week exercise and telephone follow-up program. *J Am Geriatr Soc.* Mar 2009;57(3):395-402.

[12]Jack BW, Chetty VK, Anthony D, et al. A reengineered hospital discharge program to decrease rehospitalization: a randomized trial. *Ann Intern Med.* Feb 3 2009;150(3):178-187.

[13]Koehler BE, Richter KM, Youngblood L, et al. Reduction of 30-day postdischarge hospital readmission or emergency department (ED) visit rates in high-risk elderly medical patients through delivery of a targeted care bundle. *Journal of Hospital Medicine.* Apr 2009;4(4):211-218.

[14]Weiss M, Yakusheva O, Bobay K. Nurse and patient perceptions of discharge readiness in relation to postdischarge utilization. *Med Care.* May 2010;48(5):482-486.

[15]Stauffer BD, Fullerton C, Fleming N, et al. Effectiveness and cost of a transitional care program for heart failure: a prospective study with concurrent controls. *Archives of Internal Medicine.* Jul 25 2011;171(14):1238-1243.

[16]Voss R, Gardner R, Baier R, Butterfield K, Lehrman S, Gravenstein S. The care transitions intervention: translating from efficacy to effectiveness. *Archives of Internal Medicine.* Jul 25 2011;171(14):1232-1237.

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

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

☐Clinical Practice Guideline recommendation (with evidence review)

☐US Preventive Services Task Force Recommendation

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

☐Other

Source of Systematic Review: • **Title** • **Author** • **Date** • **Citation, including page number** • **URL**	
Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	
Provide all other grades and definitions from the evidence grading system	
Grade assigned to the **recommendation** with definition of the grade	
Provide all other grades and definitions from the recommendation grading system	
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	
Estimates of benefit and consistency across studies	
What harms were identified?	
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	

1a.4 OTHER SOURCE OF EVIDENCE

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

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

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

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

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

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

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form
NQF_2515_CABG_Readmission_NQF_Evidence_Attachment_01-11-17_v1.0.docx

1b. Performance Gap

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

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

1b.1. **Briefly explain the rationale for this measure** *(e.g., the benefits or improvements in quality envisioned by use of this measure)*
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 readmission 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 readmission is a priority area for outcome measure development, as it is an outcome that is likely attributable to care processes and is an important outcome for patients. Measuring and reporting readmission rates will inform healthcare providers and facilities about opportunities to improve care, strengthen incentives for quality improvement, and ultimately improve the quality of care received by Medicare patients. The measure will also provide patients with information that could guide their choices, as well as increase transparency for consumers.

1b.2. **Provide performance scores on the measure as specified (current and over time) at the specified level of analysis.** *(This is required for endorsement maintenance. Include mean, std dev, min, max, interquartile range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
Variation in readmission rates indicates opportunity for improvement. We conducted analyses using a sample of January 1, 2009 to September 30, 2011 Medicare claims data (n=151,443 admissions from 1,195 hospitals) and reported hospital-level RSRRs having a mean of 16.8% (SD=0.02) and a range of 12.0% - 23.1%. The median RSRR is 16.8% (25th and 75th percentiles are 15.6% and 17.9%, respectively). The distribution of RSRRs across hospitals is shown below:

	RSRR(%)
Maximum	23.1
90%	19.2
75%	17.9
Median	16.8
25%	15.6
10%	14.6

Minimum 12.0

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

In 2007, the Medicare Payment Advisory Committee (MedPAC) published a report to Congress in which it identified the seven conditions associated with the most costly potentially preventable readmissions in the U.S. Among these seven, CABG ranked as having the highest potentially preventable readmission rate within 15 days following discharge (13.5%) as well as the second highest average Medicare payment per readmission ($8,136) (MedPAC 2007). The annual cost to Medicare for potentially preventable CABG readmissions was estimated at $151 million.

Variation in readmission rates indicates opportunity for improvement. Applying the measure to 2009 Medicare claims data; the range in hospital-level RSRRs is 13.3% to 21.3%.

High readmission rates and wide variation in these rates suggest that there is room for improvement. Reducing readmissions after CABG surgery has been identified as a target for quality measurement. An all-cause readmission measure for patients who undergo CABG surgery will provide hospitals with an incentive to reduce readmissions through prevention and/or early recognition and treatment of postoperative complications, and improved coordination of peri-operative care and discharge planning. Finally, CABG surgery has been identified as a potential applicable condition for use in the Affordable Care Act's Hospital Readmission Reduction Program (Office of the Legislative Counsel 2010).

References:

Medicare Payment Advisory Committee (MedPAC). Report to the Congress: Promoting Greater Efficiency in Medicare, 2007.
Office of the Legislative Counsel. Compilation of Patient Protection and Affordable Care Act 2010:6.

1b.4. Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability. *(This is required for endorsement maintenance. 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 subcriterion on improvement (4b.1) under Usability and Use.*

We conducted analyses to explore disparities in hospitals' performance on the CABG readmission measure by race and socioeconomic status (SES).

Race:

We used the Medicare Provider Analysis and Review (MEDPAR) File for 2008-2010 to calculate the percentage of African-American patients at each hospital, using all patients admitted to each hospital. We examined hospital-level RSRRs across hospitals which were grouped by decile of percentage of African-American patients for whom they cared (hospitals in the lowest decile had <0.9% African-American patients and those in the highest decile had >14% African-American patients). There was an increase in median RSRRs by decile (0.5% increase between lowest to highest) as well as a broader range of RSRRs as the proportion of African-American patients increased. The distributions for the RSRRs overlapped, and many hospitals caring for the highest percentage of African-American patients performed well on the measure. The median (range) weighted RSRR for hospitals with the highest proportion of African-American patients was 17.3% (12.5%-22.4%) compared with 16.8% (12.7%-21.9%) for hospitals with the lowest proportion of African-American patients. On the CABG readmission measure, overall the hospitals with the most African-American patients perform slightly worse than hospitals with the fewest African-American patients, but the two groups show a similar range of performance, indicating that both groups can perform well on the measures.

Note: In table below, %AA = proportion of African American patients

Decile	#Hospitals	%AA(min)	%AA(max)	RSRR(median)	RSRR(min)	RSRR(max)
.	1,197	0	100.0	16.85%	12.53%	22.36%
3	358	0	0.86	16.76%	12.66%	21.88%
4	121	0.86	1.67	16.71%	12.85%	21.09%
5	120	1.68	2.53	16.74%	13.30%	21.11%
6	119	2.53	3.72	17.21%	13.80%	21.53%
7	119	3.73	5.73	16.71%	14.11%	21.93%
8	121	5.75	8.74	16.93%	13.58%	22.07%
9	120	8.75	13.91	17.21%	14.04%	21.77%
10	119	13.96	100.0	17.34%	12.53%	22.36%

SES:
We determined a SES level for each hospital, by calculating the percentage of patients dually enrolled in both Medicare and Medicaid for each hospital, using all patients admitted to each hospital. We grouped hospital into deciles by percentage of Medicaid beneficiaries and examined hospital-level RSRRs across deciles (hospitals in the lowest decile had <3% Medicaid beneficiaries and those in the highest decile had >29% Medicaid beneficiaries). There were increases in median RSRRs across deciles (0.6% increase between lowest to highest). The median (range) weighted RSRR was 16.8% (13.3%-21.9%) for hospitals in the lowest (fewest Medicaid beneficiaries) and 17.4% (14.1%-21.1%) for the highest (most Medicaid beneficiaries) deciles. The distributions for the RSRRs overlapped and the distribution for those hospitals caring for the highest proportion of Medicaid beneficiaries was narrower than for those caring for the fewest Medicaid patients, with the worst hospital in the highest decile (most Medicaid beneficiaries) performing better on the measure than the worst hospital in the lowest decile (fewest Medicaid beneficiaries). Many hospitals in the highest decile performed well on the measure. Overall, the hospitals with the most Medicaid beneficiaries perform slightly worse than hospitals with the fewest Medicaid beneficiaries, but the two groups show a similar range of performance, indicating that both groups can perform well on the measures.

Decile	#Hospitals	%Medicaid(min)	%Medicaid(max)	RSRR(median)	RSRR(min)	RSRR(max)
.	1,197	0	100.0	16.85%	12.53%	22.36%
1	119	0	3.26	16.85%	13.30%	21.88%
2	119	3.27	5.15	16.40%	12.66%	21.10%
3	116	5.17	6.65	16.53%	13.48%	20.20%
4	125	6.67	7.86	16.85%	13.62%	21.93%
5	119	7.87	9.23	16.58%	13.34%	22.07%
6	121	9.26	11.1	16.75%	12.53%	21.68%
7	118	11.18	13.61	16.97%	13.81%	20.79%
8	121	13.64	18.55	17.37%	13.35%	22.36%
9	120	18.56	29.36	16.94%	12.76%	21.77%
10	119	29.41	100.0	17.44%	14.11%	21.11%

1b.5. **If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.**
N/A

1c. High Priority (previously referred to as High Impact)

The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. **Demonstrated high priority aspect of healthcare**
Affects large numbers, A leading cause of morbidity/mortality, Frequently performed procedure, High resource use, Patient/societal consequences of poor quality, Severity of illness
1c.2. **If Other:**

1c.3. **Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.**
CABG is a priority area for outcomes measure development because it is a common procedure associated with considerable morbidity, mortality, and health care spending. In 2007, there were 114,028 hospitalizations for CABG surgery and 137,721 hospitalizations for combined surgeries for CABG and valve procedures ("CABG plus valve" surgeries) in the U.S. (Drye et al., 2009).

Readmission rates following CABG surgery are high and vary across hospitals. The CABG unadjusted mean hospital readmission rate calculated in the January 2009-September 2011 dataset of Medicare FFS patients undergoing isolated CABG surgery is 17.7% and ranges from 0-100% with a median of 16.8% (25th and 75th percentiles are 13.1% and 20.8%, respectively). The variation persists after risk adjustment. The mean RSRR in January 2009-September 2011 data is 16.8% with a range from 12.0%-23.1%. The median risk-standardized rate is 16.8% (25th and 75th percentiles are 15.6% and 17.9%, respectively). Similarly, published data also demonstrate variation in readmission rates. The average 30-day all-cause, hospital-level readmission rate was 16.5% and ranged from

8.3% to 21.1% among patients who underwent CABG surgery in New York between January 1, 2005 and November 30, 2007 (Hannan et al., 2011). Among patients readmitted within 30 days, 87.3% of readmissions were for reasons related to CABG surgery, with a 30-day rate of readmissions due to complications of CABG surgery of 14.4%. Patients readmitted within 30 days also experienced a 2.8% in-hospital mortality rate during their readmission(s), three-fold higher than the 30-day mortality rate for patients without readmissions (Hannan et al., 2011).

1c.4. **Citations for data demonstrating high priority provided in 1a.3**

Drye E, Krumholz H, Vellanky S, Wang Y. Probing New Conditions and Procedures for New Measure Development: Yale New Haven Health Systems Corporation; Center for Outcomes Research and Evaluation.; 2009:1-7.

Hannan EL, Zhong Y, Lahey SJ, et al. 30-day readmissions after coronary artery bypass graft surgery in New York State. JACC Cardiovasc Interv. 2011;4(5):569-576.

1c.5. **If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful.** *(Describe how and from whom their input was obtained.)*

N/A. This measure is not a PRO-PM.

2. Reliability and Validity—Scientific Acceptability of Measure Properties

Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. ***Measures must be judged to meet the subcriteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria.***

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

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

Cardiovascular, Cardiovascular : Coronary Artery Disease, Surgery : Cardiac Surgery

De.6. **Cross Cutting Areas** *(check all the areas that apply):*

«crosscutting_area»

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

N/A

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

This is not an eMeasure **Attachment:**

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

Attachment **Attachment:** NQF_2515_CABG_Readmission_Data_Dictionary_01-11-17_v1.0.xlsx

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.

N/A

S.4. **Numerator Statement** *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome)*

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

The outcome for this measure is 30-day all-cause readmission. We define all-cause readmission as an unplanned inpatient admission for any cause within 30 days after the date of discharge from the index admission for patients 18 years and older who were

discharged from the hospital after undergoing isolated CABG surgery. If a patient has one or more unplanned admissions (for any reason) within 30 days after discharge from the index admission, only one is counted as a readmission.

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
Numerator time window: We define the time period for readmission as within 30 days from the date of discharge of the index CABG procedure hospitalization.

Denominator time window: This measure was developed using claims data from the calendar years 2008, 2009, and 2010. The time window can be specified from one to three years. Currently, the measure is publicly reported with three years of index admissions.

S.6. **Numerator Details** *(All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
IF an OUTCOME MEASURE, describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
This is an all-cause readmission measure and therefore any readmission within 30 days of discharge from the index hospitalization (hereafter, referred to as discharge date) is included in the measure unless that readmission is deemed a "planned" readmission. The outcome is attributed to the hospital that provided the index CABG procedure.

Planned Readmission Definition
Planned readmissions are scheduled admissions for elective procedures or for planned care such as chemotherapy or rehabilitation. Because planned readmissions are not necessarily a signal of quality of care, we chose to exclude planned readmissions from being considered as an outcome in this readmission measure. Although clinical experts agree that planned readmissions are rare after CABG, they likely do occur. Therefore, to identify these planned readmissions we have adapted and applied an algorithm originally created to identify planned readmissions for a hospital-wide (i.e., not condition-specific) readmission measure. This algorithm underwent two rounds of public comment, a validation study using data from a medical record review, and was finalized based upon technical input of 17 surgeons nominated by 9 surgical societies as well as 10 other expert surgeons.

In brief, the algorithm identifies a short list of always planned readmissions (those where the principal discharge diagnosis is major organ transplant, obstetrical delivery, or maintenance chemotherapy) as well as those readmissions with a potentially planned procedure (e.g., total hip replacement) AND a non-acute principle discharge diagnosis code. For example, a readmission for colon resection is considered planned if the principal diagnosis is colon cancer but unplanned if the principal diagnosis is abdominal pain, as this might represent a complication of the CABG procedure or hospitalization. Readmissions that included potentially planned procedures with acute diagnoses or procedures that might represent specific complications of CABG, such as PTCA or repeat CABG are not excluded from the measure outcome as they are not considered planned in this measure. Readmissions are considered planned if any of the following occurs during the readmission:

1. A procedure is performed that is in one of the procedure categories that are always planned regardless of diagnosis;
2. The principal diagnosis is in one of the diagnosis categories that are always planned; or,
3. A procedure is performed that is in one of the potentially planned procedure categories and the principal diagnosis is not in the list of acute discharge diagnoses.

Only the first readmission following an index hospital stay is counted in the numerator of this measure. If a patient has two or more readmissions within 30 days of discharge from the index hospital stay, only the first will be considered an outcome of interest; the second or later readmissions are not counted in the outcome.

Full detail, including lists of procedures and diagnoses, are included in the Measure Methodology Report in the attached appendix.

It should be noted that this approach differs from that adopted by STS for their registry-based measure, in which all 30-day readmissions were considered to be unplanned.

Outcome Attribution
Attribution of the outcome in situations where a patient has multiple contiguous admissions, at least one of which involves an index CABG procedure (i.e., the patient is either transferred into the hospital that performs the index CABG or is transferred out to another hospital following the index CABG) is as follows:

- 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 readmission outcome is attributed to the first hospital performing the index CABG procedure and the 30-day window starts with the date of discharge from the final hospital in the chain.

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 readmission risk even among transferred patients.

- 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 readmission outcome is attributed to the second hospital performing the index CABG procedure and the 30-day window starts with the date of discharge from the final hospital in the chain.

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

-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 readmission outcome is attributed to the first hospital performing the index (first) CABG procedure and the 30-day window starts with the date of discharge from the final hospital in the chain.

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 readmission risk even among transferred patients.

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
This claims-based measure can be used in either of two patient cohorts: (1) patients aged 65 years or older or (2) patients aged 18 years or older. We have tested the measure in both age groups.

The cohort includes admissions for patients a) who receive a qualifying isolated CABG procedure and b) with a complete claims history for the 12 months prior to admission. For simplicity of implementation and as testing demonstrated, closely correlated patient-level and hospital-level results using models with or without age interaction terms, the only recommended modification to the measure for application to all-payer data sets is replacement of the "Age-65" variable with a fully continuous age variable.

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Elderly, Populations at Risk

S.9. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, specific data collection items/responses , code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
The index cohort includes admissions for patients aged 18 years or older who received a qualifying "isolated" CABG procedure (CABG procedure without other concurrent major cardiac procedure such as a valve replacement). All patients in the cohort are alive at discharge (i.e., no in-hospital death). The measure was developed in a cohort of patients 65 years and older who were enrolled in Medicare FFS and admitted to non-federal hospitals. To be included in the Medicare FFS cohort, patients had to have a qualifying isolated CABG procedure AND had to be continuously enrolled in Medicare Fee-for-Service (FFS) one year prior to the first day of the index hospitalization and through 30 days post-discharge.

This cohort is defined using the ICD-9 and ICD-10 Clinical Modification procedure codes identified in Medicare Part A Inpatient claims data. The ICD-10 specifications are attached in the Data Dictionary. ICD-9 and ICD-10 procedure codes that indicate a patient has undergone a NON-isolated CABG procedure (CABG surgeries that occur concomitantly with procedures that elevate patients' readmission risk) and thus does not meet criteria for inclusion in the measure cohort are listed in the attached Data Dictionary.

ICD-9-CM codes that define the cohort:
36.10 - Aortocoronary bypass for heart revascularization, not otherwise specified
36.11 - (Aorto) coronary bypass of one coronary artery
36.12 - (Aorto coronary bypass of two coronary arteries
36.13 - (Aorto) coronary bypass of three coronary arteries
36.14 - (Aorto) coronary bypass of four or more coronary arteries
36.15 - Single internal mammary- coronary artery bypass
36.16 - Double internal mammary- coronary artery bypass
36.17 - Abdominal- coronary artery bypass

36.19 - Other bypass anastomosis for heart revascularization

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
In order to create a clinically coherent population for risk adjustment and in accordance with existing NQF-approved CABG measures and clinical expert opinion, the measure is intended to capture isolated CABG patients (i.e., patients undergoing CABG procedures without concomitant valve or other major cardiac or vascular procedures).

For all cohorts, hospitalizations are excluded if they meet any of the following criteria, for admissions:

1. Without at least 30 days post-discharge enrollment in FFS Medicare
2. Discharged against medical advice (AMA)
3. Admissions for subsequent qualifying CABG procedures during the measurement period

S.11. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
In order to create a clinically coherent population for risk adjustment and in accordance with existing NQF-approved CABG measures and clinical expert opinion, the measure is intended to capture isolated CABG patients (i.e., patients undergoing CABG procedures without concomitant valve or other major cardiac or vascular procedures).

For all cohorts, hospitalizations are excluded if they meet any of the following criteria:

1. Without at least 30 days post-discharge enrollment in FFS Medicare
Rationale: The 30-day readmission outcome cannot be assessed in this group since claims data are used to determine whether a patient was readmitted.

2. Discharged against medical advice (AMA)
Rationale: Providers did not have the opportunity to deliver full care and prepare the patient for discharge.

3. Admissions for subsequent qualifying CABG procedures during 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.

S.12. **Stratification Details/Variables** *(All information required to stratify the measure results including the stratification variables, definitions, 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 with at S.2b)*
N/A

S.13. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
Statistical risk model
If other:

S.14. **Identify the statistical risk model method and variables** *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
Our approach to risk adjustment is tailored to and appropriate for a publicly reported outcome measure, as articulated in the American Heart Association (AHA) Scientific Statement, "Standards for Statistical Models Used for Public Reporting of Health Outcomes" (Krumholz et al., 2006).

The measure calculates readmission rates using a hierarchical logistic regression model to account for the clustering of patients within hospitals while risk-adjusting for differences in patient case-mix. We modeled the log-odds of readmission within 30 days of discharge from an index CABG admission as a function of patient demographic and clinical characteristics, and a random hospital-specific intercept. This strategy accounts for within-hospital correlation of the observed outcomes, and models the assumption that underlying differences in quality among the health care groups being evaluated lead to systematic differences in outcomes.

Methodology for calculation of risk-standardized rates is noted below in the calculation algorithm section (S.18).
Variables are patient-level risk-adjustors that are expected to be predictive of readmission, based on empirical analysis, prior literature, and clinical judgment, including age and indicators of comorbidity and disease severity. For each patient, covariates are obtained from Medicare claims extending 12 months prior to and including the index admission. The model adjusts for case differences based on the clinical status of the patient at the time of admission. We use condition categories (CCs), which are clinically meaningful groupings of more than 15,000 ICD-9-CM diagnosis codes. A map showing the assignment of ICD-9 codes to CCs can be found in the attached data dictionary. We do not risk-adjust for CCs that are possible adverse events of care and that are only recorded in the index admission. In addition, only comorbidities that convey information about the patient at that time or in the 12-months prior, and not complications that arise during the course of the hospitalization are included in the risk-adjustment. The risk adjustment model includes 26 variables:

Demographics
Mean age minus 65 (SD)
Male (%)

Comorbidities
History of Coronary Artery Bypass Graft (CABG) or valve surgery (ICD-9 diagnosis codes: V42.2, V43.3, V45.81, 414.02, 414.03, 414.04, 414.05, 414.06, 414.07, 996.02, 996.03; ICD-9 procedure code: 39.61)
Cardiogenic shock (ICD-9 diagnosis code 785.51)
Chronic Obstructive Pulmonary Disease (COPD) (CC 108)
Cancer; metastatic cancer and acute leukemia (CC 7-12)
Diabetes mellitus (DM) or DM complications (CC 15-19, 119-120)
Protein-calorie malnutrition (CC 21)
Disorders of fluid/electrolyte/acid-base (CC 22-23)
Other endocrine/metabolic/nutritional disorders (CC 24)
Severe hematological disorders (CC 44)
Dementia or other specified brain disorders (CC 49-50)
Major psychiatric disorders (CC 54-56)
Hemiplegia, paraplegia, paralysis, functional disability (CC 67-69, 100-102, 177-178)
Polyneuropathy (CC 71)
Congestive heart failure (CC 80)
Specified arrhythmias and other heart rhythm disorders (CC 92-93)
Stroke (CC 95-96)
Cerebrovascular disease (CC 97-99, 103)
Vascular or circulatory disease (CC 104-106)
Fibrosis of lung or other chronic lung disorders (CC 109)
Pneumonia (CC 111-113)
Other lung disorders (CC 115)
Dialysis status (CC 130)
Renal failure (CC 131)

Please see the attached Data Dictionary for the ICD-10/V22-defined risk variables.
Risk model coefficients to estimate each patient's probability for the outcome:
SAS procedure PROC GLIMMIX fits the statistical model to calculate the risk-adjusted coefficients and hospital-specific effects as listed in the attached Data Dictionary. For random effect, the between-hospital variance is 0.04 (standard error 0.01) for the model using 2009 full year dataset.
Reference:
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.

S.15. **Detailed risk model specifications** *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.
Available in attached Excel or csv file at S.2b

S.15a. **Detailed risk model specifications** (*if not provided in excel or csv file at S.2b*)
N/A

S.16. **Type of score:**
Rate/proportion
If other:

S.17. **Interpretation of Score** *(Classifies interpretation of score according to whether better quality is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score)*
Better quality = Lower score

S.18. **Calculation Algorithm/Measure Logic** *(Describe the calculation of the measure score as an ordered sequence of steps including identifying the target population; exclusions; cases meeting the target process, condition, event, or outcome; aggregating data; risk adjustment; etc.)*
We calculate hospital-specific risk-standardized readmission rates (RSRRs). These rates are obtained as the ratio of predicted to expected readmissions, multiplied by the national unadjusted rate. The expected number of readmissions in each hospital is estimated using its patient mix and the average hospital-specific intercept. The predicted number of readmissions in each hospital is estimated given the same patient mix but the hospital-specific intercept. Operationally, the expected number of readmissions for each hospital is obtained by regressing the risk factors on the 30-day readmission using all hospitals in our sample, applying the subsequent estimated regression coefficients to the patient characteristics observed in the hospital, adding the average of the hospital-specific intercepts, summing over all patients in the hospital, and then transforming to get a count. This is a form of indirect standardization. The predicted hospital outcome is the number of expected readmissions in the "specific" hospital and not at a reference hospital. Operationally this is accomplished by estimating a hospital-specific intercept that represents baseline readmission risk within the hospital, applying the estimated regression coefficients to the patient characteristics in the hospital, summing over all patients in the hospital, and then transforming to get a count. To assess hospital performance in any given year, we re-estimate the model coefficients using that year's data.

Please see the calculation algorithm attachment for more details.

S.19. **Calculation Algorithm/Measure Logic Diagram URL or Attachment** *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
Available in attached appendix at A.1

S.20. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
N/A. This measure is not based on a sample or survey.

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
N/A. This measure is not based on a sample or survey.

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
Missing values are rare among variables used from claims data in this measure.

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Claims (Only)

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
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).

The American Community Survey (2008-2012): The American Community Survey data is collected annually and an aggregated 5-years data was used to calculate the AHRQ socioeconomic status (SES) composite index score.

Data sources for the all-payer testing: For our analyses to examine use in all-payer data, we used all-payer data from California. California is a diverse state, and, with more than 37 million residents, California represents 12% of the US population. We used the California Patient Discharge Data, a large, linked database of patient hospital admissions. In 2006, there were approximately 3 million adult discharges from more than 450 non-Federal acute care hospitals. Records are linked by a unique patient identification number, allowing us to determine patient history from previous hospitalizations and to evaluate rates of both readmission and mortality (via linking with California vital statistics records).

Using all-payer data from California, we performed analyses to determine whether the HF readmission measure can be applied to all adult patients, including not only FFS Medicare patients aged 65 years or over, but also non-FFS Medicare patients aged 18-64 years at the time of admission.

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.

S.25. **Data Source or Collection Instrument** *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No data collection instrument provided

S.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Facility

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Hospital, Hospital : Acute Care Facility
If other:

S.28. **COMPOSITE Performance Measure** - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*
N/A. This measure is not a composite performance measure.

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
NQF_2515_CABG_Readmission_NQF_Testing_Attachment_01-11-17_v1.0.docx

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b2-2b7)

Measure Number (*if previously endorsed*)**:** Click here to enter NQF number
Measure Title: Hospital 30-day, all-cause, unplanned, risk-standardized readmission rate (RSRR) following coronary artery bypass graft (CABG) surgery
Date of Submission: 1/11/2017
Type of Measure:

☒ Outcome (*including PRO-PM*)	☐ Composite – ***STOP – use composite testing form***

☐Intermediate Clinical Outcome	☐ Cost/resource
☐ Process	☐ Efficiency
☐ Structure	

Instructions

- 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.
- **For all measures, sections 1, 2a2, 2b2, 2b3, and 2b5 must be completed.**
- **For outcome and resource use measures**, section **2b4** also must be completed.
- If specified for **multiple data sources/sets of specificaitons** (e.g., claims and EHRs), section **2b6** also must be completed.
- Respond to all questions as instructed with answers immediately following the question. All information on testing to demonstrate meeting the subcriteria for reliability (2a2) and validity (2b2-2b6) must be in this form. An appendix for *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 20 pages (*incuding questions/instructions;* minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.
- For information on the most updated guidance on how to address sociodemographic variables and testing in this form refer to the release notes for version 6.6 of the Measure Testing Attachment.

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

2a2. Reliability testing [10] 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 **PRO-PMs and composite performance measures**, reliability should be demonstrated for the computed performance score.

2b2. Validity testing [11] 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 **PRO-PMs and composite performance measures**, validity should be demonstrated for the computed performance score.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [12]

AND

If patient preference (e.g., informed decisionmaking) 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). [13]

2b4. 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 sociodemographic 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.

2b5. 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.

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b7. For **eMeasures, composites, and PRO-PMs** (or other measures susceptible to missing data), 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 nonresponders) and how the specified handling of missing data minimizes bias.

Notes

10. 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).

11. 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.

12. 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.

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

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

15. 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.

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

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

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

Measure Specified to Use Data From: (***must be consistent with data sources entered in S.23***)	**Measure Tested with Data From:**
☐ abstracted from paper record	☒ abstracted from paper record
☒ administrative claims	☒ administrative claims
☐ clinical database/registry	☐ clinical database/registry
☐ abstracted from electronic health record	☐ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☐ other: Click here to describe	☒ other: Census Data/American Community Survey

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

The datasets used for testing included Medicare Parts A and B claims, Society of Thoracic Surgeons (STS) Adult Cardiac Surgery Database, California Patient Discharge Data, as well as the Medicare Enrollment Database (EDB). Additionally, census data were used to assess socioeconomic factors and race (dual eligibility and African American race variables obtained through enrollment data; Agency for Healthcare Research and Quality [AHRQ] socioeconomic status [SES] index score obtained through census data). The dataset used varies by testing type; see Section 1.7 for details.

1.3. What are the dates of the data used in testing?

We used data from January 1, 2008 throughSeptember 30, 2011 for most measure testing. We used data from calendar year 2006 for testing the measure's risk model in an all-payer (rather than Medicare FFS only) sample. For the specific dates used by the type of testing performed, see Section 1.7.

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

Measure Specified to Measure Performance of: (***must be consistent with levels entered in item S.26***)	**Measure Tested at Level of:**
☐ individual clinician	☐ individual clinician
☐ group/practice	☐ group/practice
☒ hospital/facility/agency	☒ hospital/facility/agency
☐ health plan	☐ health plan
☐ other: Click here to describe	☐ other: Click here to describe

1.5. How many and which measured entities were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of measured entities included in the*

analysis (e.g., size, location, type); if a sample was used, describe how entities were selected for inclusion in the sample)

For this measure, hospitals are the measured entities. All non-federal, acute inpatient US hospitals (including territories) that admitted Medicare FFS beneficiaries over the age of 65 for a CABG procedure are included. Between January 1, 2009 and September 30, 2011, there were 1,195 hospitals with a qualifying admission for a CABG procedure. The number of measured entities (hospitals) varies by testing type; see Section 1.7 for details.

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

The number of admissions/patients varies by testing type; see Section 1.7 for details.

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

The datasets, dates, number of measured entities and number of admissions used in each type of testing are as follows:

<u>For measure development and testing</u>
Dataset 1
To develop and validate the adequacy of the measure's statistical model, we used a combined sample of data from Medicare Part A inpatient and outpatient claims, Part B claims, and Medicare Enrollment Database (EDB) data from calendar years 2008, 2009, and 2010.
The 2008 cohort included 62,811 admissions from 1,163 hospitals
The 2009 cohort included 58,676 admissions from 1,160 hospitals
The 2010 cohort included 54,404 admissions from 1,164 hospitals.

<u>For reliability testing (Section 2a2)</u>
Medicare inpatient and outpatient claims across the 2008-2010 years of data were combined and used to test reliability of the measure. We used a combined 2008-2010 sample, randomly split it into two approximately equal subsets of patients, and calculated the RSRR for each hospital for each sample. There were 175,891 admissions in the combined three-year sample, with 87,872 admissions in one randomly selected sample and 88,019 admissions in the other randomly selected sample**.**

<u>For measures score validity testing (Section 2b2)</u>
We assessed face validity of the measaure score using a Technical Expert Panel

<u>For validiation of the measure's risk model (Section 2b2)</u>:
Dataset 1 combined with hospital-level measure results from the Society of Thoracic Surgeons (STS) readmission measure
Measure development and testing included a registry-based clinical validation study using the Society of Thoracic Surgeons (STS) Adult Cardiac Surgery Database. To validate the claims codes used to identify an isolated CABG cohort, we derived our study population from all inpatient claims for Medicare fee-for-service (FFS) patients who had an ICD-9-CM procedural code for CABG (36.1x) in any position during calendar years 2008-2010. After eliminating patients not meeting inclusion criteria for an isolated CABG procedure and applying exclusions, the final validation study population consisted of 207,656 index CABG admissions (average age of 73.9 years, 68.8% male) from 1,014 hospitals.

For testing of measure exclusions (Section 2b3)
Dataset 2 (Combined claims dataset from January 1, 2009 through September 30, 2011): Medicare Part A Inpatient and Outpatient and Part B Outpatient claims.
For the age 65+ model, we used all isolated CABG admissions between January 1, 2009 and September 30, 2011 in Medicare FFS data. The final cohort included 150,900 admissions (average age of 73.9 years, 69.0% male) from 1,195 hospitals.

For testing of measure risk adjustment (Section 2b4)

Dataset 1

For Sub-section 2b4.11. Optional Additional Testing for Risk Adjustment

Dataset 3 (all payer dataset, section 2b4.11): California Patient Discharge Data in addition to CMS Medicare FFS data for patients in California hospitals
We also applied the model to all-payer data from California. The analytic sample included 14,635 isolated CABG cases aged 18 and older (average age of 65.9 years, 75.0% male) in the 2006 California Patient Discharge Data. When used in all-payer data, only admission claims data are used for risk adjustment, as the hospital discharge databases do not have outpatient claims.

Testing to identify meaningful differences in performance (Section 2b5)
Dataset 1

For testing of sociodemographic factors in risk models (Section 2b4.4b)

Dataset 4 (2015 public reporting dataset): This dataset included Medicare FFS claims for all index admissions for a qualifying CABG procedure from July 1, 2011 through June 30, 2014.

Number of Admissions: N=137,958 cases matched to FFS Medicare claims
Number of Measured Entities: 1,199

Dataset 5 (The American Community Survey [ACS]): The American Community Survey, 2008-2012

We examined disparities in performance according to the proportion of patients in each hospital who were of African-American race and the proportion who were dual eligible for both Medicare and Medicaid insurances. We also used the AHRQ SES index score to study the association between performance measures and socioeconomic status.

Data Elements

• African-American race and dual eligible status (i.e., enrolled in both Medicare and Medicaid) patient-level data are obtained from CMS enrollment data (**Dataset 4**).
• Validated AHRQ SES index score is a composite of 7 different variables found in the census data

1.8 What were the patient-level sociodemographic (SDS) variables that were available and analyzed in the data or sample used? For example, patient-reported data (e.g., income, education, language), proxy variables when SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate).

Sociodemographic status incorporates socioeconomic variables as well as race into a more concise term. However, given the fact that socioeconomic risk factors are distinct from race and should be interpreted differently, we have decided to keep “socioeconomic status” and “race” as separate terms.

We selected socioeconomic status (SES) and race variables to analyze after reviewing the literature and examining available national data sources. There is a large body of literature linking various SES factors and African-American race to worse health status and higher readmission risk (Blum et al., 2014; Eapen et al. 2015; Gilman et al., 2014; Hu et al., 2014; Joynt and Jha, 2013). Income, education, and occupational level are the most commonly examined variables. However, while literature directly examining how different SES factors or race might influence the likelihood of older, insured, Medicare patients of being readmitted within 30 days of an admission for heart failure is more limited, studies indicate an association between SES/race and increased risk of heart failure readmission (Foraker et al., 2011; Kind et al., 2014; Vivo et al., 2014; Joynt, Orav, and Jha 2011; Lindenauer et al., 2013; Allen et al., 2012; Regalbuto et al., 2014; Calvillo-King et al., 2013; McHugh, Carthon, and Kang 2010;). The causal pathways for SES and race variable selection are described below in Section 2b4.3.

The SES and race variables used for analysis were:

- Dual eligible status (**Dataset 4**)
- African-American race (**Dataset 4**)
- AHRQ-validated SES index score (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) (**Dataset 5**)

In selecting variables, our intent was to be responsive to the NQF guidelines for measure developers in the context of the SDS Trial Period. Our approach has been to examine all patient-level indicators of both SES and race/ethnicity that are reliably available for all Medicare beneficiaries and linkable to claims data and to select those that are most valid.

Previous studies examining the validity of data on patients' race and ethnicity collected by CMS have shown that only the data identifying African-American beneficiaries have adequate sensitivity and specificity to be applied broadly in research or measures of quality. While using this variable is not ideal because it groups all non-African-American beneficiaries together, it is currently the only race variable available on all beneficiaries across the nation that is linkable to claims data.

We similarly 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 both our race and the dual-eligible variables, there is a body of literature demonstrating differential health care and health outcomes among beneficiaries indicating that these variables, while not ideal, also allow us to examine some of the pathways of interest.

Finally, we selected the AHRQ-validated SES index score because it is a well-validated and widely-used variable that describes the average socioeconomic status of people living in defined geographic areas. 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. Currently, the individual data elements used to calculate the score are available at the 9-digit census block group zip code level. However, in this submission, we present analysis using the 5-digit level. We have performed these analyses with SES data attributed at the census block level, the most granular level possible, for several other readmission measures and have presented those results to this committee at past in-person meetings.

References:

Allen LA, Smoyer Tomic KE, Smith DM, Wilson KL, Agodoa I. Rates and predictors of 30-day readmission among commercially insured and Medicaid-enrolled patients hospitalized with systolic heart failure. *Circulation. Heart failure.* 2012;5(6):672-679.

Blum AB, Egorova NN, Sosunov EA, et al. Impact of socioeconomic status measures on hospital profiling in New York City. Circulation. Cardiovascular quality and outcomes. May 2014; 7(3):391-397.

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.

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

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.

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 Aff (Millwood). Aug 2014; 33(8):1314-22.

Hu J, Gonsahn MD, Nerenz DR. Socioeconomic status and readmissions: evidence from an urban teaching hospital. Health affairs (Project Hope). 2014; 33(5):778-785.

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

Joynt, K. E., E. J. Orav and A. K. Jha. "Thirty-Day Readmission Rates for Medicare Beneficiaries by Race and Site of Care." JAMA 305, no. 7 (2011): 675-81.

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.

Lindenauer, P. K., T. Lagu, M. B. Rothberg, J. Avrunin, P. S. Pekow, Y. Wang and H. M. Krumholz. "Income Inequality and 30 Day Outcomes after Acute Myocardial Infarction, Heart Failure, and Pneumonia: Retrospective Cohort Study." Bmj 346, (2013): f521.

McHugh MD, Carthon JM, Kang XL. Medicare readmissions policies and racial and ethnic health disparities: a cautionary tale. *Policy, politics & nursing practice.* 2010;11(4):309-316.

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.

Vivo, R. P., S. R. Krim, L. Liang, M. Neely, A. F. Hernandez, Z. J. Eapen, E. D. Peterson, D. L. Bhatt, P. A. Heidenreich, C. W. Yancy and G. C. Fonarow. "Short- and Long-Term Rehospitalization and Mortality for Heart Failure in 4 Racial/Ethnic Populations." J Am Heart Assoc 3, no. 5 (2014): e001134.

2a2. RELIABILITY TESTING

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

2a2.1. What level of reliability testing was conducted? (*may be one or both levels*)

☒ **Critical data elements used in the measure** (*e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements*)

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

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

Data Element Reliability

In constructing the measure, we aim to utilize only those data elements from the claims that have both face validity and reliability. We avoid the use of fields that are thought to be coded inconsistently across hospitals or providers. Specifically, we use fields that are consequential for payment and which are audited. We identify such variables through empiric analyses and our understanding of CMS auditing and billing policies and seek to avoid variables which do not meet this standard. For example, "discharge disposition" is a variable in Medicare claims data that is not thought to be a reliable variable for identifying a transfer between two acute care facilities. Thus, we derive a variable using admission and discharge dates as a surrogate for "discharge disposition" to identify hospital admissions involving transfers. This allows us to identify these admissions using variables in the claims data which have greater reliability than the "discharge disposition" variable.

In addition, CMS has in place several hospital auditing programs used to assess overall claims code accuracy, to ensure appropriate billing, and for overpayment recoupment. CMS routinely conducts data analysis to identify potential problem areas and detect fraud, and audits important data fields used in our measures, including diagnosis and procedure codes and other elements that are consequential to payment.

Finally, we assess the reliability of the data elements by comparing model variable frequencies and odds ratios from logistic regression models across the most recent three years of data (**Dataset 1**).

Measure Score 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 comparing 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 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), and assessed the values according to conventional standards (Landis and Koch, 1977). Specifically, we used **Dataset 1** split sample and calculated the RSRR for each hospital for each sample. The agreement of the two RSRRs 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, compared with 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.

References:

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

Landis J, Koch G. The measurement of observer agreement for categorical data. Biometrics 1977;33:159-174.

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.

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

Data element reliability results (**Dataset 1**)

Overall, risk factor frequencies changed very little across the three-year period, and there were no notable differences in the odds ratios across years of data. (Results are included in the Measure Methodology Report in the attached appendix).

Measure Score Reliability Results

There were 175,891 admissions in the combined three-year sample (from **Dataset 1**), with 87,872 admissions in one of the randomly selected samples and 88,019 admissions in the other randomly selected sample, each mutually exclusive of the other. The agreement between the two RSRRs for each hospital was 0.331, which according to the conventional interpretation is "fair".[1] The intra-class correlation coefficient is based on a split sample of 3 years of data, resulting in a volume of patients in each sample equivalent to only 1.5 years of data, whereas the measure is likely to be publicly reported with a full three years of data. Based on our experiences with similar measures using split samples, from 4 years of data (and a sample volume equivalent to 2 years), the intra-class correlation coefficient would be higher and likely in the "moderate" range.

References:

Landis J, Koch G. The measurement of observer agreement for categorical data, Biometrics 1977;33:159-174.

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

The stability of the risk factor frequencies and odds ratios indicates data elements are reliable. Additionally, the ICC score demonstrates fair agreement across samples, indication that the measure score is reliable.

2b2. VALIDITY TESTING

2b2.1. What level of validity testing was conducted? (*may be one or both levels*)

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

☒ **Performance measure score**

☐ **Empirical validity testing**

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

2b2.2. For each level of testing checked above, describe the method of validity testing and what it tests (*describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used)*

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

Validity of Claims-Based Measures

Our team has demonstrated for a number of prior measures the validity of claims-based measures for profiling hospitals by comparing either the measure results or individual data elements against medical records. CMS validated the eight NQF-endorsed measures currently in public reporting (AMI, heart failure, COPD, and pneumonia mortality and readmission) with models that used chart-abstracted data for risk adjustment. Specifically, claims model validation was conducted by building comparable models using abstracted medical chart data for risk adjustment for heart failure patients (National Heart Failure data), AMI patients (Cooperative Cardiovascular Project data) and pneumonia patients (National Pneumonia Project dataset). When both models were applied to the same patient population, the hospital risk-standardized rates estimated using the claims-based risk adjustment models had a high level of agreement with the results based on the medical record model, thus supporting the use of the claims-based models for public reporting. Our group has reported these findings in the peer-reviewed literature.[1-6]

Validity Indicated by Established Measure Development Guidelines

We developed this measure in consultation with national guidelines for publicly reported outcomes measures, with outside experts, and with the public. The measure is consistent with the technical approach to outcomes measurement set forth in NQF guidance for outcomes measures[7] (National Quality Forum, 2010), CMS Measure Management System (MMS) guidance, and the guidance articulated in the American Heart Association scientific statement, "Standards for Statistical Models Used for Public Reporting of Health Outcomes".[8]

Validity as Assessed by External Groups

Throughout measure development, we obtained expert and stakeholder input via three mechanisms: regular discussions with an advisory working group, a national TEP, and a 30-day public comment period in order to increase transparency and to gain broader input into the measure.

The working group was comprised of two cardiothoracic surgeons with expertise in quality measure development, one of whom was the lead for the development of the STS registry-based CABG readmission measure. In addition, two members of the claims-based measure development team served on the working group for the STS CABG readmission measure. Through frequent (weekly or more frequent) conference calls, all aspects of measure development were discussed among the two measure developers, including the cohort definitions, outcome attribution, and risk-adjustment. The collaboration allowed real-time harmonization of the measures throughout the entire measure development process. The working group meetings addressed key issues surrounding measure development, including detailed discussions regarding the appropriate cohort for inclusion in the measure. The working group provided a forum for focused expert review and discussion of technical issues during measure development prior to consideration by the broader, combined TEP, which was convened to address all three CABG outcomes measures under development (the two claims-based readmission and mortality measures as well as the registry-based readmission measure). This allowed for continuation of the close collaboration between measure developers achieved earlier in measure development.

In addition to the working group, and in alignment with the CMS Measure Management System, we convened a TEP to provide input and feedback during measure development from a group of recognized experts in relevant fields. To convene the TEP, we released a public call for nominations and selected individuals to represent a

range of perspectives including clinicians, consumers, and purchasers, as well as individuals with experience in quality improvement, performance measurement, and health care disparities. We held three structured TEP conference calls consisting of presentation of key issues, our proposed approach, and relevant data, followed by open discussion among TEP members. We made minor modifications to the measure cohort (i.e., excluding additional concomitant non-cardiac procedures from the cohort such as lung resection and mastectomy), and risk-adjustment variables (i.e., including a history of prior CABG surgery in the risk adjustment) based on TEP feedback on the measures.

Following completion of the model, we solicited public comment on the measure through the CMS site link https://www.CMS.gov/MMS/17_CallforPublicComment.asp. The public comments were then posted publicly for 30 days.

Face Validity as Determined by TEP

To systematically assess face validity, we surveyed the Technical Expert Panel 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 readmission rates obtained from the readmission measure as specified will provide an accurate reflection of quality."

Validity of the Measures Cohort and Risk-Adjustment Model as Assessed by Registry Data Validation

In collaboration with the Society of Thoracic Surgeons (STS), we performed a validation study of this measure using the national STS Adult Cardiac Surgery Database, including the following:

Validation of the administrative isolated CABG cohort

Validation of the administrative isolated CABG cohort consisted of matching, using probabilistic matching at the patient and hospital level, the administrative CABG cohort for the administrative readmission measure detailed in this application to the measure cohort for the proposed STS registry data-based CABG readmission measure. Non-matching patients were identified as either claims only patients (i.e., the administrative cohort defined them as isolated CABG patients while the STS registry did not) or registry only patients (i.e., the administrative cohort defined them as non-isolated CABG patients while the registry defined them as isolated CABG patients). This information was then used to further harmonize the administrative readmission cohort inclusion/exclusion criteria and codes to align as much as possible with the registry definition of isolated CABG procedures.

Validation of the administrative risk adjustment model

Validation of the administrative risk adjustment model consisted of comparing the hospital-level RSRRs and performance category assigned by the administrative CABG readmission measure detailed in this application in the matched cohort of CABG patients to the RSRRs calculated and performance category assigned by the STS clinical data-based CABG readmission measure (also in the matched cohort and using identical methods for defining the outcome and performance categorization). For each of the two measures, RSRRs were estimated in a hierarchical logistic regression model with hospital-specific random intercept parameters. Methods of estimation were identical to the currently publicly reported CMS mortality and readmission measures for Acute Myocardial Infarction, Heart Failure and Pneumonia. A bootstrapping algorithm was used to construct a 95% interval estimate for each RSRR. To complete this analysis, we categorized hospitals into three performance groups -- "Better", "Same" and "Worse" than the national rate -- according to the methodology used for the currently publicly reported CMS mortality and readmission measures. We classified a hospital as performing "Better than the national rate" if the 95% interval estimate for that hospital was entirely below the overall aggregate readmission rate for all hospitals, "Worse than the national rate" if the estimate for that hospital was entirely above the overall aggregate readmission rate, and "Same as the national rate" if the estimate included the overall aggregate readmission rate.

Statement of Intent and Process of Conversion

This application includes ICD-10 codes that correspond to the ICD-9 codes included in our measure specifications. The goal of conversion to ICD-10 was to convert this measure to a new code set, fully consistent with the intent of the original measure. ICD-10 codes were initially identified using 2016 GEM mapping software. We then enlisted the help of clinicians with expertise in relevant areas to select and evaluate which ICD-10 codes map to the ICD-9 codes currently in use for this measure. We examined this ICD-10 code set in a 6-month sample of ICD-10-coded claims submitted by hospitals after October 1,2016. The ICD-10-based specifications are attached in field the Data Dictionary.

References:

1. Krumholz HM, Wang Y, Mattera JA, Wang Y-F, Han LF, Ingber MJ, Roman S, Normand SL. An administrative claims model suitable for profiling hospital performance based on 30-day mortality rates among patients with an acute myocardial infarction. Circulation. 2006 Apr 4;113(13):1683-92.
2. 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.
3. Krumholz HM, Wang Y, Mattera JA, Wang Y-F, Han LF, Ingber MJ, Roman S, Normand SL. An administrative claims model suitable for profiling hospital performance based on 30-day mortality rates among patients with heart failure. Circulation. 2006 Apr 4;113(13):1693-701.
4. 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.
5. Bratzler DW, Normand SL, Wang Y, O'Donnell WJ, Metersky M, Han LF, Rapp MT, Krumholz HM. An administrative claims model for profiling hospital 30-day mortality rates for pneumonia patients. Public Library of Science One. 2011 Apr 12;6(4):e17401.
6. 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.
7. 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.
8. 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.* 2006;113(3):456-462.

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

Validity as Assessed by External Groups

Fourteen TEP members responded to the survey question as follows: Moderately Disagreed (2), Somewhat Disagreed (2), Somewhat Agreed (4), Moderately Agreed (5), and Strongly Agreed (1). Hence, 71% of TEP members agreed (43% moderately or strongly agreed) that the measure will provide an accurate reflection of quality.

Registry Data Validation

Validation of administrative isolated CABG cohort

The cohort validation demonstrated an overall agreement rate of 96.5% (200,475 of 207,656 matched patients were designated as isolated or non-isolated CABG patients by both measure cohort definitions). Among the

4,720 patients identified as isolated CABG by the claims measure but not by the registry measure, 37% were due to expected causes (i.e., the fact that the registry measure excludes all MAZE procedures while the claims measure excludes only open MAZE procedures).The remaining 2,976 patients identified as isolated CABG by the claims measure but not by the registry measure and the 2,461 patients identified as isolated CABG patients by the registry measure but not by the claims measure were due to inconsistencies that could not clearly be attributed to inaccuracies in the claims-based definition of the isolated CABG cohort. For example, among a proportion of patients, the patient had a code for an aortic valve replacement but the registry data did not show that this procedure was performed. Alternatively, the registry data indicated an aortic valve procedure was performed but there was no corresponding claims code for this procedure. Such inconsistencies could be due to coding errors in the claims data, abstraction errors in the registry data, or may be due to inconsistencies in the probabilistic matching process used to create a matched set of patients for the validation. An additional reason that patients might be identified as isolated CABG patients by the registry measure but not by the claims measure is that the CABG procedure occurred on a separate day within the index admission than the valve or other procedure that excluded the patient from the claims-based isolated CABG cohort. Only two of 286 such discrepant aortic valve procedures could be attributed to procedures occurring on different days during the index admission. Among the discrepant patients, the non-CABG-related ICD-9 procedure codes represented only nonspecific ancillary procedures to CABG surgery, such as code 39.61 "Extracorporeal circulation auxiliary to open heart surgery" and could not be used to further increase the precision of the administrative claims-based isolated CABG cohort definition. The level of agreement for this measure was significantly higher than prior studies comparing administrative definitions of isolated CABG to registry data.[4]

Validation of administrative risk adjustment model

Both the claims-based and registry-based measures displayed similar distributions in hospital RSRRs following CABG and the median hospital RSRR differed by only 0.1% point (16.7% and 16.8% for registry-based and claims-based measures, respectively).

The comparison of the risk adjustment performance of the administrative and clinical models in a matched set of patients produced an overall agreement of 97% (807 of 829 hospitals had concurrent performance categorization) and the correlation was between 0.92 and 0.96, depending upon the statistic used. No hospitals were rated as performing Worse than the national rate by the claims-based measure and Better than the national rate by the registry-based measure (or vice versa). Among 14 hospitals rated Better than the national rate by the registry-based measure, 8 were rated No different than the national rate by the claims-based measure and among 9 hospitals rated Better than the national rate by the claims-based measure, 3 were rated No different than the national rate by the registry-based measure. Among 14 hospitals rated Worse by the registry model, 6 were rated No different than the national rate by the claims model and among 13 programs rated Worse by the claims model, 5 were rated No different than the national rate by the registry model.

Overall, 63 of 829 hospitals (7.6%) had greater than a 1% absolute difference in RSRR calculated by the claims-based versus registry-based measures. However, of these 63, only 8 hospitals actually changed performance category.

References:

1. 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.

2. 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.

3. 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.

4. 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.

2b2.4. What is your interpretation of the results in terms of demonstrating validity? (i*.e., what do the results mean and what are the norms for the test conducted?*)

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.

Registry Data Validation

Validation of administrative isolated CABG cohort

The results of the cohort validation using the companion CABG readmission measure and the national STS Adult Cardiac Surgery database did not suggest the need for any changes to the cohort definition. The claims-based cohort definition of isolated CABG was nearly identical to that assigned by registry data. The level of agreement greatly exceeded that of previous efforts for CABG.[1] The discrepant patients were either due to expected differences due to the respective measure cohort definitions (e.g., MAZE procedures, which are handled differently in the two measures) or to reasons that cannot be clearly ascribed to errors or inadequacies in the claims-based definition.

Validation of administrative risk adjustment model

The risk-adjustment validation provides evidence of the claims-based measure's scientific soundness.
The risk-adjustment validation produced a substantial correlation of RSRRs between the two measures in a matched cohort of patients, with an intraclass correlation coefficient of 0.92. When hospitals were categorized as "Better", "Worse" or "No different" than the national rate, over 97% (807 of 829) of hospitals in the matched cohort were categorized identically by the two measures (the vast majority were considered "No different than the national rate" by either measure). Twenty-two hospitals were assigned to an outlier category ("Better" or "Worse") by one measure but not by the other; however, no hospital was rated as "Better" by one measure and "Worse" by the other (or vice versa). The individual RSRRs estimated by the claims-based measure for the 22 hospitals with discordant performance categorization all fell within the 95% interval estimates for the RSRR estimated by the registry-based measure.

Even where there is disagreement in the performance category, the measures profile hospitals similarly -- all better performing hospitals (those with either their claims- or registry-based interval estimates below the national rate) have RSRRs for both measures well below the national readmission rate); conversely, the worse performing hospitals (those with either their claims- or registry-based interval estimates above the national rate) have RSRRs for both measures well above the national rate. The differences in the results could have implications for a small number of individual hospitals if these classifications are used for assigning payments or penalties. The implications of the differences will depend on the specifics of the public reporting and/or payment programs using the results and merit careful consideration.

Finally it is important to note that the validation of the claims-based measure risk adjustment is only generalizable to STS hospitals. Because the STS registry does not capture all patients in all hospitals, and because non-STS hospitals do not represent a random sample of hospitals, the validation results only provide information as to the performance of the claims-based measure in STS hospitals. The risk model used in the claims-based measure uses information from both STS and non-STS hospitals in selecting and estimating the

impact of risk variables, but, as the STS model is only developed in STS hospitals, this validation work cannot assess the performance of the claims-based measure in other hospitals. However, the STS registry represents the largest and most comprehensive dataset available for this type of validation.

Reference:
1. 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.

2b3. EXCLUSIONS ANALYSIS

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

2b3.1. Describe the method of testing exclusions and what it tests (*describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used*)

Exclusions were those determined by expert input to be clinically relevant. These exclusions are consistent with similar NQF-endorsed readmission measures. Rationales for the exclusions are detailed in Denominator Exclusions section (S.10). To ascertain impact of exclusions on the cohort, we examined overall frequencies and proportions of the total cohort excluded for exclusions that are not data requirements (such that, without the data, measure calculation would not be possible), or have minimal impact on the measure due to very low frequency.

2b3.2. What were the statistical results from testing exclusions? (*include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores*)

For the purposes of tabulation, exclusions are performed sequentially. Thus, a hospital stay that would be excluded based on multiple criteria is counted in the first criterion only. Among 1,195 hospitals with at least 25 index stays in January 2009 – September 2011 (**Dataset 2**):

Exclusion	N	%	Distribution across hospitals
1. Hospital stays in which patients leave hospital against medical advice (AMA)	40	0.03%	n/a (low impact)
2. Hospital stays for patients without at least 30 days post-discharge information	494	0.33%	n/a (data-related exclusion)
3. Subsequent hospital stays for patients with additional CABG procedure admissions within 30 days	9	0.01%	n/a (low impact)

These exclusions represent 0.37% of the initial cohort (n=151,443). We do not report frequency of distribution of exclusions across measured entities due to the minimal impact of the exclusions on the measure cohort.

2b3.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results? (*i.e., the value outweighs the burden of increased data collection and analysis. Note: **If patient preference is an exclusion**, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion*)

The exclusions listed above were based on clinical input or are required for the determination of the outcome. Exclusion 1 is needed because, while very few patients are discharged AMA, the exclusion is needed for acceptability of the measure to hospitals. Exclusions 2 and 3 are necessary for valid calculation of the measure.

2b4. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES

If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b5.

2b4.1. What method of controlling for differences in case mix is used?

☐ **No risk adjustment or stratification**

☒ **Statistical risk model with 26 risk factors**

☐ **Stratification by** Click here to enter number of categories **risk categories**

☐ **Other,** Click here to enter description

2b4.1.1 If using a statistical risk model, provide detailed risk model specifications, including the risk model method, risk factors, coefficients, equations, codes with descriptors, and definitions.

See data dictionary and item 2b4.3.

2b4.2. If an outcome or resource use component measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities.

N/A

2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or sociodemographic factors) used in the statistical risk model or for stratification by risk (*e.g., potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10; correlation of x or higher; patient factors should be present at the start of care*)

Our approach to risk adjustment was tailored to, and appropriate for, a publicly reported outcome measure, as articulated in the American Heart Association (AHA) Scientific Statement, "Standards for Statistical Models Used for Public Reporting of Health Outcomes" (Krumholz et al. 2006).

The measure employs a hierarchical logistic regression model (a form of hierarchical generalized linear model [HGLM]) to create a hospital-level 30-day RSRR. This approach to modeling appropriately accounts for the structure of the data (patients clustered within hospitals), the underlying risk due to patients' comorbidities, and sample size at a given hospital when estimating hospital readmission rates. In brief, the approach simultaneously models two levels (patient and hospital) to account for the variance in patient outcomes within and between hospitals (Normand and Shahian et al. 2007). At the patient level, each model adjusts the log-odds of readmission within 30-days of admission for age, sex, selected clinical covariates and a hospital-specific intercept. The second level models the hospital-specific intercepts as arising from a normal distribution. The hospital intercept, or hospital-specific effect, represents the hospital contribution to the risk of readmission, after accounting for patient risk and sample size, and can be inferred as a measure of quality. The hospital-specific intercepts are given a distribution in order to account for the clustering (non-independence) of patients within the same hospital. If there were no differences among hospitals, then after adjusting for patient risk, the hospital intercepts should be identical across all hospitals.

Clinical Factors

Candidate and Final Risk-adjustment Variables: The original measure was developed using Medicare FFS claims data. Candidate variables were patient-level risk-adjustors that are expected to be predictive of readmission, based on empirical analysis, prior literature, and clinical judgment, including demographic factors (age, sex) and indicators of comorbidity and disease severity. For each patient, covariates were obtained from Medicare claims extending 12 months prior to and including the index admission. The model adjusted for case differences based on the clinical status of the patient at the time of admission. We used condition categories (CCs), which are clinically meaningful groupings of more than 15,000 ICD-9-CM diagnosis codes. We did not risk-adjust for CCs that were possible adverse events of care and that were only recorded in the index admission. In addition, only comorbidities that conveyed information about the patient at that time or in the 12-months prior, and not complications that arose during the course of the hospitalization were included in the risk adjustment.

The original ICD-9-based risk adjustment variables were:

Demographics

Mean age minus 65 (SD)

Male (%)

Comorbdities

History of Coronary Artery Bypass Graft (CABG) or valve surgery (ICD-9 diagnosis codes: V42.2, V43.3, V45.81, 414.02, 414.03, 414.04, 414.05, 414.06, 414.07, 996.02, 996.03; ICD-9 procedure code: 39.61)

Cardiogenic shock (ICD-9 diagnosis code 785.51)

Chronic Obstructive Pulmonary Disease (COPD) (CC 108)

Cancer; metastatic cancer and acute leukemia (CC 7-12)

Diabetes mellitus (DM) or DM complications (CC 15-19, 119-120)

Protein-calorie malnutrition (CC 21)

Disorders of fluid/electrolyte/acid-base (CC 22-23)

Other endocrine/metabolic/nutritional disorders (CC 24)

Severe hematological disorders (CC 44)

Dementia or other specified brain disorders (CC 49-50)

Major psychiatric disorders (CC 54-56)

Hemiplegia, paraplegia, paralysis, functional disability (CC 67-69, 100-102, 177-178)

Polyneuropathy (CC 71)

Congestive heart failure (CC 80)

Specified arrhythmias and other heart rhythm disorders (CC 92-93)

Stroke (CC 95-96)

Cerebrovascular disease (CC 97-99, 103)

Vascular or circulatory disease (CC 104-106)

Fibrosis of lung or other chronic lung disorders (CC 109)

Pneumonia (CC 111-113)

Other lung disorders (CC 115)

Dialysis status (CC 130)

Renal failure (CC 131)

Please see the attached Data Dictionary for the ICD-10/V22-defined risk variables.

Socioeconomic Status (SES) Factors and Race

We selected variables representing socioeconomic status (SES) factors and race 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 SES and race may influence 30-day readmission.

Our conceptualization of the pathways by which patient SES or race affects 30-day readmission is informed by the literature.

Literature Review of Socioeconomic Status (SES) and Race Variables and CABG Readmission

To examine the relationship between SES and race variables and hospital 30-day, all-cause, risk-standardized readmission rate (RSRR) 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 SES or race and CABG readmission. Nine studies were initially reviewed, and seven studies were excluded from full-text review based on the above criteria. Studies have been limited, and those that have been conducted have used travel distance and living alone as variables (Chou, Deily, and Li 2014; Murphy et al. 2008), with results being too limited to indicate a consistent effect.

Causal Pathways for Socioeconomic Status (SES) and Race Variable Selection

Although some recent literature evaluates the relationship between patient SES or race and the readmission outcome, few studies directly address causal pathways or examine the role of the hospital in these pathways. 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 readmission. The SES factors that have been examined in the readmission 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 range from the self-reported or documented race or ethnicity of the patient to the patient's income or education level (Eapen et al., 2015; Hu et al., 2014). Neighborhood/community-level variables use information from sources such as the American Community Survey (ACS) 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 Agency for Healthcare Research and Quality (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 SES risk factors influence the risk of readmission 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 socioeconomic status (SES) factors or race 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 SES 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**.**

In addition to SES risk factors, studies have shown that worse health status is more prevalent among African-American patients compared with white patients. The association between race and worse health is in part mediated by the association between race and SES risk factors such as poverty or disparate access to care associated with poverty or neighborhood. The association is also mediated through bias in healthcare as well as other facets of society.

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 contribute to increased risk of readmission following hospitalization (Jha et al., 2011; Reames et al., 2014). Similarly African-American patients have been shown to have less access to high quality facilities compared with white patients (Skinner et al., 2005).

3. **Differential care within a hospital**. The third major pathway by which SES factors or race may contribute to readmission risk is that patients may not receive equivalent care within a facility. For example, African-American patients have been shown to experience differential, lower quality, or discriminatory care within a given facility (Trivedi et al., 2014). Alternatively, patients with SES 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 SES on readmission risk outside of hospital quality and health status**. Some SES risk factors, such as income or wealth, may affect the likelihood of readmission 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 SES and race variables were considered:

• Dual eligible status

• African American race

• AHRQ SES index

We assessed the relationship between the SES variables and race 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.

One concern with including SES or race factors in a model is that their effect may be at either the patient or the hospital level. For example, low SES may increase the risk of readmission because patients of low SES have an individual higher risk (patient-level effect) or because patients of low SES are more often admitted to hospitals with higher overall readmission rates (hospital-level effect). Thus, as an additional step, we performed a decomposition analysis to assess the independent effects of the SES and race variables at the patient level and the hospital level. If, for example, all the elevated risk of readmission for patients of low SES was due to lower quality/higher readmission risk in hospitals with more patients of low SES, then a significant hospital-level effect would be expected with little-to-no patient-level effect. However, if the increased readmission risk was solely related to higher risk for patients of low SES regardless of hospital effect, then a significant patient-level effect would be expected and a significant hospital-level effect would not be expected.

Specifically, we decomposed each of the SES and race variables as follows: Let X_{ij} be a binary indicator of the SES or race status of the i^{th} patient at the j^{th} hospital, and $\mathbf{X}_j$ the percent of patients at hospital j with $X_{ij} = 1$. Then we rewrote $X_{ij} = (X_{ij} - \mathbf{X_j}) + \mathbf{X}_j \equiv X_{patient} + X_{hospital}$. The first variable, $X_{patient}$, represents the effect of the risk factor at the patient level (sometimes called the "within" hospital effect), and the second, $X_{hospital}$, represents the effect at the hospital level (sometimes called the "between" hospital effect). By including both of these in the same model, we can assess whether these are independent effects, or whether only one of these effects contributes. This analysis allows us to simultaneously estimate the independent effects of: 1) hospitals with higher or lower proportions of low SES patients or African-American patients on the readmission rate of an average patient; and 2) a patient's SES or race on their own readmission rates when seen at an average hospital.

It is very important to note, however, that even in the presence of a significant patient-level effect and absence of a significant hospital-level effect, the increased risk could be partly or entirely due to the quality of care patients receive in the hospital. For example, biased or differential care provided within a hospital to low-income patients as compared to high-income patients would exert its impact at the level of individual patients, and therefore be a patient-level effect. It is also important to note that the patient-level and hospital-level coefficients cannot be quantitatively compared because the patient's SES circumstance or race in the model is binary whereas the hospitals' proportion of low SES patients or African-American patients is continuous.

References:

Allen LA, Smoyer Tomic KE, Smith DM, Wilson KL, Agodoa I. Rates and predictors of 30-day readmission among commercially insured and Medicaid-enrolled patients hospitalized with systolic heart failure. *Circulation. Heart failure.* 2012;5(6):672-679.

Aseltine RH, Jr., Yan J, Gruss CB, Wagner C, Katz M. Connecticut Hospital Readmissions Related to Chest Pain and Heart Failure: Differences by Race, Ethnicity, and Payer. *Connecticut medicine.* 2015;79(2):69-76.

Berenson J, Shih A. Higher readmissions at safety-net hospitals and potential policy solutions. *Issue brief (Commonwealth Fund).* 2012;34:1-16.

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.

Damiani G, Salvatori E, Silvestrini G, et al. Influence of socioeconomic factors on hospital readmissions for heart failure and acute myocardial infarction in patients 65 years and older: evidence from a systematic review. *Clinical interventions in aging.* 2015;10:237-245.

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.

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.

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 Aff (Millwood). Aug 2014; 33(8):1314-22.

Hu J, Gonsahn MD, Nerenz DR. Socioeconomic status and readmissions: evidence from an urban teaching hospital. Health affairs (Project Hope). 2014; 33(5):778-785.

Jha AK, Orav EJ, Epstein AM. Low-quality, high-cost hospitals, mainly in South, care for sharply higher shares of elderly black, Hispanic, and Medicaid patients. Health Affairs 2011; 30:1904-11.

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

Joynt, K. E., E. J. Orav and A. K. Jha. "Thirty-Day Readmission Rates for Medicare Beneficiaries by Race and Site of Care." JAMA 305, no. 7 (2011): 675-81.

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.

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.

Lindenauer, P. K., T. Lagu, M. B. Rothberg, J. Avrunin, P. S. Pekow, Y. Wang and H. M. Krumholz. "Income Inequality and 30 Day Outcomes after Acute Myocardial Infarction, Heart Failure, and Pneumonia: Retrospective Cohort Study." Bmj 346, (2013): f521.

McHugh MD, Carthon JM, Kang XL. Medicare readmissions policies and racial and ethnic health disparities: a cautionary tale. *Policy, politics & nursing practice.* 2010;11(4):309-316.

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.

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.

Skinner J, Chandra A, Staiger D, Lee J, McClellan M. Mortality after acute myocardial infarction in hospitals that disproportionately treat black patients. Circulation 2005; 112:2634-41.

Trivedi AN, Nsa W, Hausmann LR, et al. Quality and equity of care in U.S. hospitals. The New England journal of medicine 2014; 371:2298-308.

Vivo, R. P., S. R. Krim, L. Liang, M. Neely, A. F. Hernandez, Z. J. Eapen, E. D. Peterson, D. L. Bhatt, P. A. Heidenreich, C. W. Yancy and G. C. Fonarow. "Short- and Long-Term Rehospitalization and Mortality for Heart Failure in 4 Racial/Ethnic Populations." J Am Heart Assoc 3, no. 5 (2014): e001134.

2b4.4a. What were the statistical results of the analyses used to select risk factors?

Below is a table showing the original ICD-9-based variables in the model with associated odds ratios (OR). Please note that the current ICD-10-based risk variables are listed in the Data Dictionary.

Final Model Variables (variables meeting criteria in field 2b4.3)

Variable	**01/01/2009-09/30/2011 OR (95% CI)**
Age minus 65 (years above 65, continuous)	1.03 (1.02 – 1.03)
Male	0.77 (0.75 – 0.79)
History of prior CABG or valve surgery (ICD-9 Diagnosis Codes: V42.2, V43.3, V45.81, 414.02, 414.03, 414.04, 414.05, 414.06, 414.07, 996.02, 996.03 ; ICD-9 Procedure Codes: 39.61)	1.05 (0.99 – 1.11)
Cardiogenic shock (ICD-9 Code 785.51)	1.33 (1.24 – 1.41)
Chronic obstructive pulmonary disease (COPD) (CC 108)	1.29 (1.25 – 1.33)
Renal failure (CC 131)	1.29 (1.24 – 1.34)
Diabetes mellitus (DM) or DM complications (CC 15-20, 119-120)	1.15 (1.12 – 1.19)
Other endocrine/metabolic/nutritional disorders (CC 24)	0.85 (0.82 – 0.89)
Congestive heart failure (CC 80)	1.21 (1.17 – 1.26)
Specified arrhythmias and other heart rhythm disorders (CC 92-93)	1.12 (1.09 – 1.16)
Other lung disorders (CC 115)	1.06 (1.03 – 1.10)
Major psychiatric disorders (CC 54-56)	1.22 (1.14 – 1.30)
Vascular or circulatory disease (CC 104-106)	1.11 (1.07 – 1.14)
Disorders of fluid/electrolyte/acid-base (CC 22-23)	1.19 (1.15 – 1.24)
Pneumonia (CC 111-113)	1.16 (1.11 – 1.21)
Cerebrovascular disease (CC 97-99, 103)	0.95 (0.92 – 0.98)
Polyneuropathy (CC 71)	1.20 (1.14 – 1.26)

Variable	01/01/2009-09/30/2011 OR (95% CI)
Protein-calorie malnutrition (CC 21)	1.26 (1.18 – 1.34)
Severe hematological disorders (CC 44)	1.38 (1.23 – 1.54)
Fibrosis of lung or other chronic lung disorders (CC 109)	1.10 (1.03 – 1.17)
Decubitus ulcer or chronic skin ulcer (CC 148-149)	1.30 (1.21 – 1.39)
Dialysis status (CC 130)	1.36 (1.23 – 1.50)
Hemiplegia, paraplegia, paralysis, functional disability (CC 67-69, 100-102, 177-178)	1.12 (1.04 – 1.21)
Stroke (CC 95-96)	1.07 (1.00 – 1.14)
Dementia or other specified brain disorders (CC 49-50)	1.16 (1.09 – 1.23)
Cancer (CC 7-12)	0.99 (0.95 – 1.02)

2b4.4b. Describe the analyses and interpretation resulting in the decision to select SDS factors (e.g. prevalence of the factor across measured entities, empirical association with the outcome, contribution of unique variation in the outcome, assessment of between-unit effects and within-unit effects)

Variation in prevalence of the factor across measured entities

The prevalence of SES factors and African-American patients in the CABG cohort varies across measured entities. The median percentage of dual eligible patients is 7.1% (interquartile range [IQR]: 4.4% – 11.0%). The median percentage of African-American patients is 2.7% (IQR: 0.8% – 7.0%). The median percentage of patients with an AHRQ SES Index score equal to or below 46.0 is 18.5% (IQR: 7.5% – 37.2%).

Empirical association with the outcome (univariate)

The patient-level observed CABG readmission rate is higher for dual eligible patients, 19.53%, compared with 14.53% for all other patients. Similarly the readmission rate for patients with an AHRQ SES Index score equal to or below 46.0 was 16.10% compared with 14.57% for patients with an AHRQ SES Index score above 46.0. The readmission rate for African-American patients was also higher at 17.93% compared with 14.78% for patients of all other races.

Incremental effect of SES variables and race in a multivariable model

We then examined the strength and significance of the SES variables and race 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 modest. The c-statistic is unchanged with the addition of any of these variables into the model. Furthermore the addition of any of these variables into the model has little to no effect on hospital performance. We examined the change in hospitals' RSRRs with the addition of any of these variables. The median absolute change in hospitals' RSRRs when adding a dual eligibility indicator is 0.010% (IQR: -0.018% – 0.030%, minimum -0.316% – maximum 0.103%) with a correlation coefficient between RSRRs for each hospital with and without dual eligibility added of 0.99928. The median absolute change in hospitals' RSRRs when adding a race indicator is 0.003% (IQR: -0.003% – 0.007%, minimum -0.089% – maximum 0.018%) with a correlation coefficient between

RSRRs for each hospital with and without race added of 0.99995. The median absolute change in hospitals' RSRRs when adding an indicator for a low AHRQ SES Index score is 0.030% (IQR: -0.051% – 0.091%, minimum -1.158% – maximum 0.365%) with a correlation coefficient between RSRRs for each hospital with and without an indicator for a low AHRQ SES Index score added of 0.99205.

As an additional step, a decomposition analysis was performed. The results are described in the table below.

The patient-level and hospital-level dual eligible effects were significantly associated with CABG readmission in the decomposition analysis. If the dual eligible were used in the model to adjust for patient-level differences, then some of the differences between hospitals would also be adjusted for, potentially obscuring a signal of hospital quality.

The patient-level race and low AHRQ SES Index effects were not appreciably different from zero in the decomposition analysis, though the hospital-level race and low AHRQ SES effects were significant. If race or low AHRQ SES Index are used as risk-adjustment variables, they will primarily capture an effect of the hospital on the outcome, not the effect of intrinsic characteristics of patients or of how they are treated.

Given these findings and the complex pathways that could explain any relationship between SES or race with readmission, we did not incorporate SES variables or race into the measure.

CABG Readmission Decomposition Analysis

Parameter	Estimate (Standard Error)	P-value
Dual Eligible – Patient-Level	0.1705 (0.0269)	<.0001
Dual Eligible – Hospital-Level	0.3400 (0.1467)	0.0205
African American – Patient-Level	0.0067 (0.0347)	0.8472
African American – Hospital-Level	0.5452 (0.1403)	0.0001
AHRQ SES Index – Patient-Level	0.0357 (0.0202)	0.0777
AHRQ SES Index – Hospital-Level	0.2185 (0.0512)	<.0001

* The p-values represent the significance of the patient-level and hospital-level variables. It is important to note that the coefficients cannot be quantitatively compared because the patient-level variable is binary whereas the hospital-level variable is continuous.

2b4.5. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (*describe the steps—do not just name a method; what statistical analysis was used*)

Approach to assessing model performance

We computed three summary statistics for assessing model performance (Harrell and Shih, 2001) for the cohorts (**Dataset 1**):

Discrimination statistics:

(1) Area under the receiver operating characteristic (ROC) curve (the c-statistic (also called ROC) 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)

Reference:
1. F..E. Harrell and Y.C.T. Shih, Using full probability models to compute probabilities of actual interest to decision makers, *Int. J. Technol. Assess. Health Care* **17** (2001), pp. 17–26.

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.
If stratified, skip to 2b4.9

2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

2009 development cohort:
C-statistic = 0.62
Predictive ability (lowest decile %, highest decile %): (8.7, 29.8)

2008 validation cohort:
C-statistic = 0.63
Predictive ability (lowest decile %, highest decile %): (8.8, 30.5)

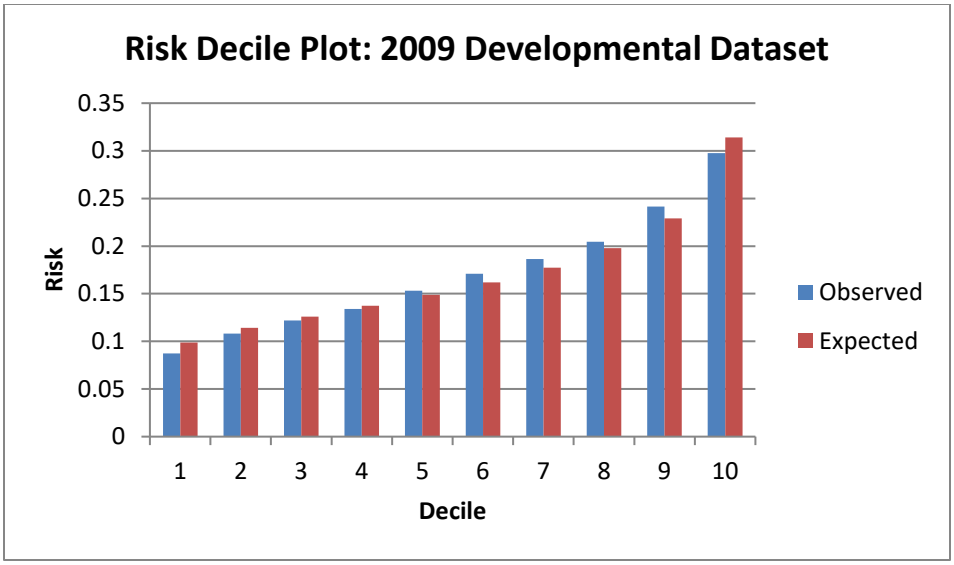
2010 validation cohort:
C-statistic = 0.63
Predictive ability (lowest decile %, highest decile %): (8.4, 30.3)

2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

2009 development cohort: Calibration (over-fitting statistics): (0, 1)
2008 validation cohort: Calibration (over-fitting statistics): (0.02, 1.01)
2010 validation cohort: Calibration (over-fitting statistics): (-0.03, 1.00)

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:

The risk decile plot is a graphical depiction of the deciles calculated to measure predictive ability. A risk decile plot for the 2009 developmental dataset, representative of risk decile plots for all other datasets, is shown below:



2b4.9. Results of Risk Stratification Analysis:

N/A

2b4.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)? (i.*e., what do the results mean and what are the norms for the test conducted*)

Discrimination Statistics
The C-statistic of 0.62 was not substantially different across datasets 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
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.

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

2b4.11. Optional Additional Testing for Risk Adjustment (*not required, but would provide additional support of adequacy of risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed*)

Application to Patients Aged 18 Years and Older
When the model was applied to all patients aged 18+ in 2006 California Patient Discharge Data, overall discrimination was good (C statistic=0.66). In addition, there was good discrimination and predictive ability in

both those aged 18-64 and those aged 65+. Moreover, the distribution of Pearson residuals was comparable across the patient subgroups. When comparing the model with and without interaction terms [between age (≥65 and <65) and individual risk factors]: (a) the reclassification analysis demonstrated 85%-95% overall agreement in patient risk categorization; (b) the C statistic was identical (0.66 in both models); and (c) hospital-level risk-standardized rates were highly correlated (ICC=0.998). Although the interaction term Older and Pneumonia was statistically significant in this analysis, the inclusion of interactions did not substantively affect either patient-level model performance or hospital-level results. Therefore, the measure can be applied to all-payer data for patients 18 years and older. For simplicity and pending further study, the only change currently recommended to the measure specifications to allow application to an all-payer, 18+ year population is transformation of the Age variable from "Age – 65" to a fully continuous age variable.

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

2b5.1. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified (*describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided related to performance gap in 1b)*

For public reporting of the measure, CMS characterizes the uncertainty associated with the RSRR by estimating the 95% interval estimate. This is similar to a 95% confidence interval but is calculated differently. If the RSRR's interval estimate does not include the national observed readmission rate (is lower or higher than the rate), then CMS is confident that the hospital's RSRR 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 RSRR 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.

2b5.2. What were the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities? (e.g., *number and percentage of entities with scores that were statistically significantly different from mean or some benchmark, different from expected; how was meaningful difference defined*)

Using the January 2009 – September 2011 cohort, unadjusted hospital-level readmission rates range from 0%-100% (25th and 75th percentile are 13.1% - 20.8%, respectively). This may be a signal of differences in the quality of care received for patients following a qualifying CABG procedure. The results of the RSRRs showed continued meaningful difference even after risk-adjustment, ranging from 12.0% - 23.1% (25th-75th percentile is 15.6% - 17.9%).

2b5.3. What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities? (i.*e., what do the results mean in terms of statistical and meaningful differences?*)

The variation in rates and number of performance outliers suggests there remain differences in 30-day all-cause readmission following a qualifying CABG procedure.

2b6. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS

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

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

2b6.1. Describe the method of testing conducted to compare performance scores for the same entities across the different data sources/specifications (*describe the steps—do not just name a method; what statistical analysis was used*)

N/A

2b6.2. What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications? (*e.g., correlation, rank order*)

N/A

2b6.3. What is your interpretation of the results in terms of the differences in performance measure scores for the same entities across the different data sources/specifications? (i.*e., what do the results mean and what are the norms for the test conducted*)

N/A

2b7. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b7.1. Describe the method of testing conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)

N/A

2b7.2. What is the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data? (*e.g., results of sensitivity analysis of the effect of various rules for missing data/nonresponse; if no empirical sensitivity analysis, identify the approaches for handling missing data that were considered and pros and cons of each*)

N/A

2b7.3. What is your interpretation of the results in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how

the specified handling of missing data minimizes bias**?** (i.*e., what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; if no empirical analysis, provide rationale for the selected approach for missing data*)

N/A

3. Feasibility

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

3a. Byproduct of Care Processes

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

3a.1. **Data Elements Generated as Byproduct of Care Processes.**
Coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims)
If other:

3b. Electronic Sources

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

3b.1. **To what extent are the specified data elements available electronically in defined fields?** *(i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields)*
ALL data elements are in defined fields in electronic claims

3b.2. **If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.**

3b.3. **If this is an eMeasure, provide a summary of the feasibility assessment in an attached file or make available at a measure-specific URL.**
No feasibility assessment **Attachment:**

3c. Data Collection Strategy

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

3c.1. **Describe what you have learned/modified as a result of testing and/or operational use of the measure regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.**
IF a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.
Administrative data are routinely collected as part of the billing process.

3c.2. **Describe any fees, licensing, or other requirements to use any aspect of the measure as specified** *(e.g., value/code set, risk model, programming code, algorithm)*.
There are no fees associated with the use of this measure.

4. Usability and Use

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

or populations.

4a. Accountability and Transparency

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

4.1. Current and Planned Use

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

Planned	Current Use (for current use provide URL)
Not in use	Public Reporting Hospital Inpatient Quality Reporting (IQR) Program http://cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/HospitalRHQDAPU.html Payment Program Hospital Readmission Reduction (HRRP) Program http://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Readmissions-Reduction-Program.html

4a.1. For each CURRENT use, checked above, provide:

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

Public Reporting
Program Name, Sponsor: Hospital Inpatient Quality Reporting (IQR) Program, Centers for Medicare and Medicaid Services (CMS)

Purpose: The Hospital Inpatient Quality Reporting (Hospital IQR) program was originally mandated by Section 501(b) of the Medicare Prescription Drug, Improvement, and Modernization Act (MMA) of 2003. This section of the MMA authorized CMS to pay hospitals that successfully report designated quality measures a higher annual update to their payment rates. Initially, the MMA provided for a 0.4 percentage point reduction in the annual market basket (the measure of inflation in costs of goods and services used by hospitals in treating Medicare patients) update for hospitals that did not successfully report. The Deficit Reduction Act of 2005 increased that reduction to 2.0 percentage points.

In addition to giving hospitals a financial incentive to report the quality of their services, the Hospital IQR program provides CMS with data to help consumers make more informed decisions about their health care. Some of the hospital quality of care information gathered through the program is available to consumers on the Hospital Compare website at: www.hospitalcompare.hhs.gov.

Geographic area and number and percentage of accountable entities and patients included:
The IQR program includes all Inpatient Prospective Payment System (IPPS) non-federal acute care hospitals and VA hospitals in the United States. The number and percentage of accountable hospitals included in the program, as well as the number of patients included in the measure, varies by reporting year. For 2015 public reporting, the RSRR was reported for 4,663 hospitals across the U.S. The final index cohort includes 925,315 admissions.

Payment Program
Program Name, Sponsor: Hospital Readmission Reduction (HRRP) Program, Centers for Medicare and Medicaid Services (CMS)

Purpose: Section 3025 of the Affordable Care Act added section 1886(q) to the Social Security Act establishing the Hospital Readmissions Reduction Program, which requires CMS to reduce payments to IPPS hospitals with excess readmissions, effective for discharges beginning on October 1, 2012. The regulations that implement this provision are in subpart I of 42 CFR part 412 (§412.150 through §412.154).

Geographic area and number and percentage of accountable entities and patients included: The HRRP program includes only Subsection (d) hospitals and hospitals located in Maryland. Subsection (d) hospital encompasses any acute care hospital located in

one of the fifty States or the District of Columbia which does not meet any of the following exclusion criteria as defined by the Social Security Act: psychiatric, rehabilitation, children's, or long-term care hospitals, and cancer specialty centers. By definition, all other hospitals are considered subsection (d) hospitals. This means that critical access hospitals, cancer hospitals, and hospitals located in U.S territories will not be included in the calculation. The number and percentage of accountable entities included in the program, as well as the number of patients included in the measure, varies by reporting year.

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

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

4b. Improvement

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

4b.1. **Progress on Improvement. (Not required for initial endorsement unless available.)**
Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (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**

There has been significant progress in 30-day RSRR for CABG. The mean RSRR decreased by over the three-year period, from 15.0% between July 2012 and June 2013 to 13.9% between July 2014 and June 2015. The median hospital RSRR in the combined three-year dataset was 14.4% (Interquartile Range [IQR] 13.8% - 15.0%).

4b.2. **If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.**
N/A

4c. Unintended Consequences

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

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**
We did not identify any unintended consequences during measure development, model testing, or re-specification. However, we are committed to monitoring this measure's use and assessing potential unintended consequences over time, such as the inappropriate shifting of care, increased patient morbidity and mortality, and other negative unintended consequences for patients.

5. Comparison to Related or Competing Measures

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

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

Yes

5.1a. **List of related or competing measures (selected from NQF-endorsed measures)**
0114 : Risk-Adjusted Postoperative Renal Failure
0115 : Risk-Adjusted Surgical Re-exploration
0119 : Risk-Adjusted Operative Mortality for CABG
0129 : Risk-Adjusted Postoperative Prolonged Intubation (Ventilation)
0130 : Risk-Adjusted Deep Sternal Wound Infection
0131 : Risk-Adjusted Stroke/Cerebrovascular Accident
0330 : Hospital 30-day, all-cause, risk-standardized readmission rate (RSRR) following heart failure (HF) hospitalization
0505 : Hospital 30-day all-cause risk-standardized readmission rate (RSRR) following acute myocardial infarction (AMI) hospitalization.
0506 : Hospital 30-day, all-cause, risk-standardized readmission rate (RSRR) following pneumonia hospitalization
1551 : Hospital-level 30-day risk-standardized readmission rate (RSRR) following elective primary total hip arthroplasty (THA) and/or total knee arthroplasty (TKA)
1789 : Hospital-Wide All-Cause Unplanned Readmission Measure (HWR)

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

5a. Harmonization

The measure specifications are harmonized with related measures;
OR
The differences in specifications are justified

5a.1. **If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s):**
Are the measure specifications completely harmonized?
Yes

5a.2. **If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.**
The proposed CABG readmission measure, which has been developed in close collaboration with STS, has a target population (i.e., isolated CABG patients) that is harmonized with the above measures to the extent possible given the differences between clinical and administrative data. The exclusions are nearly identical to the STS measures' cohort exclusions with the exception of epicardial MAZE procedures; STS excludes these procedures from the registry-based CABG readmission measure cohort because the version of registry data used for measure development did not allow them to differentiate them from open maze procedures. The age range for the proposed CABG readmission and existing NQF-endorsed STS measure cohorts differs; STS measures are specified for age 18 and over, and the proposed CABG readmission measure is currently specified for age 65 and over. However, we have performed testing in patients 18 years and over and determined the measure performs well across all adult patients and payers. The proposed CABG readmission measure is harmonized with the above measures to the extent possible given the different data sources used for development and reporting.

5b. Competing Measures

The measure is superior to competing measures (e.g., is a more valid or efficient way to measure);
OR
Multiple measures are justified.

5b.1. **If this measure conceptually addresses both the same measure focus and the same target population as NQF-endorsed measure(s):**
Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality); OR provide a rationale for the additive value of endorsing an additional measure. (Provide analyses when possible.)
There are no existing NQF-endorsed measures or other measures in current use that have the same measure focus and the same target population as this measure. However, this measure was developed concurrently with a clinical registry data-based readmission measure (Risk-adjusted readmission measure for coronary artery bypass graft (CABG)). The measure steward for the registry-based readmission measure for CABG is also CMS; STS developed the measure. Effort was taken to harmonize both the registry-based and administrative-based measures to the extent possible given the differences in data sources.

CMS developed these two "competing" measures at the same time to allow for maximum flexibility in implementation for quality improvement programs across different care settings. The STS cardiac surgery registry currently enrolls most, but not all, patients receiving CABG surgeries in the U.S. The proposed CABG readmission measure will capture all qualifying Medicare FFS patients undergoing CABG regardless of whether their hospital or surgeon participates in the STS registry.

This claims-based CABG readmission measure was developed with the goal of producing a measure with the highest scientific rigor and broadest applicability. The measure is harmonized with the above existing and proposed measures to the extent possible given the different data sources used for development and reporting.

Appendix

A.1 Supplemental materials may be provided in an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be organized in one file with a table of contents or bookmarks. If material pertains to a specific submission form number, that should be indicated. Requested information should be provided in the submission form and required attachments. There is no guarantee that supplemental materials will be reviewed.
Attachment **Attachment:** CABG_Readmission_MeasureMethodologyReport_02-01-14_Final.pdf

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** Centers for Medicare & Medicaid Services
Co.2 **Point of Contact:** Lein, Han, Lein.han@cms.hhs.gov, 410-786-0205-
Co.3 **Measure Developer if different from Measure Steward:** Yale New Haven Health Services Corporation/Center for Outcomes Research and Evaluation (YNHHSC/CORE)
Co.4 **Point of Contact:** Karen, Dorsey, karen.dorsey@yale.edu, 203-764-5700-

Additional Information

Ad.1 **Workgroup/Expert Panel involved in measure development**
Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.
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

Measure Developer/Steward Updates and Ongoing Maintenance
Ad.2 **Year the measure was first released:** 2015
Ad.3 **Month and Year of most recent revision:**
Ad.4 **What is your frequency for review/update of this measure?** N/A
Ad.5 **When is the next scheduled review/update for this measure?**

Ad.6 **Copyright statement:** N/A Ad.7 **Disclaimers:** N/A
Ad.8 **Additional Information/Comments:** N/A