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 #: 3188
De.2. **Measure Title:** 30-Day Unplanned Readmissions for Cancer Patients
Co.1.1. **Measure Steward:** Seattle Cancer Care Alliance
De.3. **Brief Description of Measure:** 30-Day Unplanned Readmissions for Cancer Patients measure is a cancer-specific measure. It provides the rate at which all adult cancer patients covered as Fee-for-Service Medicare beneficiaries have an unplanned readmission within 30 days of discharge from an acute care hospital. The unplanned readmission is defined as a subsequent inpatient admission to an acute care hospital, which occurs within 30 days of the discharge date of an eligible index admission and has an admission type of "emergency" or "urgent."
1b.1. **Developer Rationale:** For many cancer patients, readmission following hospitalization may be preventable and should be addressed to potentially lower costs and improve patient outcomes. The ADCC recognizes the need for oncology-specific efficiency measures, including unplanned readmissions because planned readmissions are often used in clinical pathways for cancer patients. In 2014, the ADCC identified C4QI's 30-Day Unplanned Readmissions for Cancer Patients measure as a potential accountability measure for the PPS-Exempt Cancer Hospitals Quality Reporting Program (PCHQR). C4QI's 21 members (11 ADCC hospitals/PCHs and 10 other academic medical centers, or AMC) have utilized this claims-based, cancer-specific unplanned readmissions measure since 2012. It is designed to reflect the unique clinical aspects of oncology and to provide a more comprehensive measurement of unplanned readmissions in cancer patients, when compared with existing measures (e.g., the HWR measure). It considers patients with an admission type of "emergency" or "urgent" within 30 days of an index admission as an unplanned readmission. It excludes readmissions for patients readmitted for chemotherapy or radiation therapy treatment or with disease progression. Using this measure, hospitals can better identify and address preventable readmissions for cancer patients.

An earlier version of this measure (NQF #2884) was reviewed by the NQF All-Cause Admissions and Readmissions Project 2015-2017 Technical Expert Panel (TEP) in June 2016. Following the recommendation of the TEP, the ADCC broadened the measure to capture readmissions of cancer patients to any short-term acute care PPS hospital and pursued additional testing of the measure using Medicare claims data (i.e., the Standard Analytical Files). This expansion produced unplanned readmissions rates of patients discharged from PCHs and readmitted to any short-term acute care hospital (defined as PCHs, short-term acute care Prospective Payment System, or PPS, hospitals, and Critical Access Hospitals, or CAH). Additionally, it provided comparative rates of unplanned readmissions of cancer patients for non-PCH short–term acute care hospitals (i.e., short-term acute care PPS hospitals and CAHs).

S.4. **Numerator Statement:** This outcome measure demonstrates the rate at which adult cancer patients have an unplanned readmissions at an acute care hospital within 30 days of discharge from an eligible index admission. The numerator includes all eligible unplanned readmissions to an acute care hospital within 30 days of the discharge date from an index admission that is included in the measure denominator. Readmissions with an admission type of "emergency" or "urgent" are considered unplanned readmissions within this measure.
Additional details are provided in S.5 Numerator Details.
S.6. **Denominator Statement:** The denominator includes inpatient admissions for all adult Fee-for-Service Medicare beneficiaries where the patient is discharged from an acute care hospital with a principal or secondary diagnosis (i.e., not admitting diagnosis) of malignant cancer within the defined measurement period.
S.8. **Denominator Exclusions:** The following index admissions are excluded from the measure denominator:
1) Less than 18 years of age;
2) Patients who died during the index admission;
3) Patients discharged AMA;
4) Patients transferred to another acute care hospital during the index admission;
5) Patients discharged with a planned readmission;
6) Patients having missing or incomplete data; and,

7) Patients not admitted to an inpatient bed.

De.1. **Measure Type:** Outcome
S.17. **Data Source:** Claims (Only)
S.20. **Level of Analysis:** Facility

IF Endorsement Maintenance – Original Endorsement Date: Most Recent Endorsement Date:

Preliminary Analysis

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.

Criteria 1: Importance to Measure and Report

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.

- As a rationale for measuring this health outcome, the developer lists several studies from peer-reviewed journals explaining that cancer is the second cause of death in the United States, with nearly 600,000 cancer-related deaths expected this year.
- Developers explain that this measure intends to reflect the unique clinical aspects of oncology patients and to yield readmission rates that may be obscured by a broader readmission measure, such as the Hospital-Wide All-Cause Unplanned Readmission Measure (HWR).
- The developer notes that there are several clinical actions that can be taken by the accountable entity to improve the outcome of 30-day readmissions. Specifically, the logic model notes that providers can ensure that patients are clinically ready for discharge with clear and appropriate follow-up care planned. These actions will help foster improved patient care, better population health, and reduce readmission risk.

Summary of prior review in All-Cause Admissions and Readmissions 2015-2017 Project

- Measure 2884, the previous version of this measure, was included in the Admissions and Readmissions 2015-2017 project.
- During the prior review of the measure, the Standing Committee recommended expanding the measure definition to include cancer readmissions all acute care hospital, and not limit to PPS-exempt cancer hospitals.
- Standing Committee Members agreed unanimously the measure met the evidence criterion.

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:**

- Updates: Seven new references added that detail unplanned readmissions for cancer patients as well as hospital-wide all-cause readmissions.

Questions for the Committee:

- *Is there at least one intervention that the provider can do 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

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

- The developers updated performance gap data using the Medicare 100% standard analytic file.
- A total of 4,975 short-term acute care hospitals (defined as PCHs, short-term acute care PPS hospitals, and CAHs) were included across 2013-2015.

	2013-2015
Mean (SD)	16.54% (8.24%)
Range	0-100%
Quartile Range	8.30%
25th percentile	12.50%
50th percentile	17.32%
75th percentile	20.80%

- These data generally represent a range of performance among hospitals

Disparities

- The developer provided descriptive statistics for several patient-level demographic characteristics including gender, age, beneficiary race code, and dual eligibility status. Developer provides frequency and percent distribution by strata of demographic categories. Readmission performance scores by strata of demographic categories is not provided by the developer in this section.
- In testing the SDS factors in the risk adjustment model, the developers noted that there was a conceptual and empirical rationale for adjustment based on dual-eligibility status. The developers note that dual-eligibility can serve as a proxy for low income status and other measures of SDS. Several studies were referenced that note that low SDS factors are a risk factor for later-state cancer diagnosis, delayed health care receipt, and higher utilization of hospital-based care.
- The patient-level observed 30-Day Unplanned Readmissions for Dual-Eligible Cancer Patients rate was 22.49%, compared with an 18.32% observed rate for all other patients.

Questions for the Committee:

- *Is there a gap in care for this area of measurement 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:

** There is direct evidence to support this outcome measure. There are numerous healthcare actions to improve this measure and improve patient care.

** Pass

** Cancer is a leading cause of death in 40 to 79 year old population. 86% of cancers are diagnosed in 50+. In 2010, cancer related health care accounted for $125 billion in health care spending. This measure which is tied to eligible index admissions and admission types that are emergency or urgent will help to move along cancer related measures which have lagged behind. There appears to be evidence that when patients are clinically ready for discharge and there is appropriate follow up care and management of co-morbidities, unnecessary readmissions in a 30 day post-discharge period can be reduced. Could there be greater clarity on meaning of emergency and urgent in this context?

** Pass

** Acceptable rationale and evidence exist to support measure.

** The evidence strongly relates to the health outcome measured. Additionally, healthcare actions such as discharge planning are identified and supported in the rationale.

** Updated by developer to include additional references

** There is sufficient evidence to support the measure with more than one intervention identified that could potentially reduce readmissions in cancer patients.

** The desired outcome identified by the measure developer is lower healthcare costs related the treatment of cancer patients. By encouraging fewer unplanned readmissions for patients with principle or secondary diagnoses of malignant cancer, the developer cites several studies that maintain that fewer unplanned readmissions will reduce overall, system-wide costs associated with cancer treatment.

** Yes

** Provided updated evidence
Clinical measures can be done prior to discharge to impact 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:

** Used a large sample of hospital to evaluate the gap. Significant variability and a range of performance among hospitals. Dual eligibility served as a proxy for SDS and there is evidence to support this. There is a significant performance gap which demonstrates the importance of the measure.

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

** This measure developed by ADCC was reviewed in 2016 by a NQF TEP. It was recommended that it be expanded to include all short term acute care hospitals and that additional testing be done using Medicare claims data. This was undertaken.

There was considerable attention to risk factors. Readmission rates for dual eligibles was 22.49% compared to 18.32% for other patients with considerable variability. Dual Eligible status was regarded as a proxy for low income and other SDS measures and included in the risk adjustment model. (Low SDS is a factor for later stage cancer diagnosis, delayed health care and higher utilization of hospital-based care.) Apparently, while there is some evidence that racial minorities have higher readmission rates, the studies are conflicting and it is difficult to discern what is attribute to patient's race v. site of care.

** Acceptable- high

** Based on provided data, there appears to be a performance gap. Re: disparities, developers elected to use dual

eligibility as a proxy for SDS and report on a higher rate of readmissions for dual eligible pts.

** Performance data was provided which demonstrates variability in performance sufficient to warrant a performance measure. Performance by strata of demographic categories was not provided.

** Large performance gap offered by developer as compared to non-cancer patients. Dual eligible pts may be diagnosed later in disease and represent delay in treatment as opposed to preventable readmissions.

** Performance data was provided for a large data set with race, sex, age and dual eligibility status identified. Would like to find a way of determining and incorporating patient-level SDS in the data set.

** The developer cites studies that identify costs/waste in the treatment of cancer patients that could be avoided via fewer unplanned readmissions.

** Yes, a gap exists

** Higher readmission for dual -eligible 22.49% as compared to 18.32% for all other patients

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):

Medicare Administrative Claims
Medicare 100% Standard Analytic File

Specifications:

- This outcome measure demonstrates the rate at which adult cancer patients have unplanned readmissions within 30 days of discharge from an eligible index admission.
- The numerator includes all eligible unplanned readmissions to any short-term acute care hospital—defined as admission to a PPS-Exempt Cancer Hospital (PCH), a short-term acute care Prospective Payment (PPS) hospital, or Critical Access Hospital (CAH)—within 30 days of the discharge date from an index admission that is included in the measure denominator. Readmissions with an admission type (UB-04 Uniform Bill Locator 14) of "emergency = 1" or "urgent = 2" are considered unplanned readmissions within this measure. Readmissions for patients with progression of disease (using a principal diagnosis of metastatic disease as a proxy) and for patients with planned admissions for treatment (defined as a principal diagnosis of chemotherapy or radiation therapy) are excluded from the measure numerator.
- The denominator includes inpatient admissions for all adult Fee-for-Service Medicare beneficiaries where the patient is discharged from a short-term acute care hospital (PCH, short-term acute care PPS hospital, or CAH) with a principal or secondary diagnosis (i.e., not admitting diagnosis) of malignant cancer within the defined measurement period.
- The following index admissions are excluded from the measure denominator:
  1) Less than 18 years of age;
  2) Patients who died during the index admission;
  3) Patients discharged AMA;
  4) Patients transferred to another acute care hospital during the index admission;
  5) Patients discharged with a planned readmission;
  6) Patients having missing or incomplete data; and,
  7) Patients not admitted to an inpatient bed.
- The following numerator exclusions are noted:
  - Readmissions for patients with progression of disease, defined as Primary Claim Diagnosis Code of metastatic disease (ICD-9-CM range: 196-198.89, 209.70 - 209.79; ICD-10-CM range: C77.0 – C79.9, C7B.0-C7B.8)
    - Developer Rationale: A primary (or principal) diagnosis of metastatic disease serves as a proxy for disease progression. Readmissions for conditions or symptoms associated with disease progression are not reflective of poor clinical care but, rather, advanced disease.
  - Readmissions for patients with planned admissions for treatment, defined as Primary Claim Diagnosis Code of chemotherapy or radiation encounter (ICD-9-CM range: V58.00-V58.12; ICD-10-CM range: Z51.00 – Z51.12).
    - Developer Rationale: Readmissions are expected and planned for some patients who require additional cancer treatment in the inpatient setting. These readmissions reflects high-quality care that is focused on patient safety and are reliably distinguishable in claims data.
- The measure is specified for a facility level of analysis and the hospital setting.
- The statistical risk adjustment model includes 11 risk factors with 15 values.
  - The developers use a logistic regression to estimate the probability of an unplanned readmission.
  - The probability of unplanned readmission is summed over the index admissions for each hospital to calculate the expected unplanned readmission rate.
  - The developers sum the actual or observed unplanned readmissions for each hospital and calculated the ratio of observed unplanned readmissions to expected unplanned readmissions for each hospital.
  - Each hospital's ratio was then multiplied by the national or standard unplanned readmissions rate to generate the risk-adjusted 30-Day Unplanned Readmissions for Cancer Patients rate (see formula below). Lower risk-adjusted rates (observed/expected ratios) are interpreted as better quality while higher risk-adjusted rates (observed/expected ratios) indicate poorer quality.

$$Risk - Adjusted\ Rate = \frac{observed\ rate}{expected\ rate}$$

Questions for the Committee:

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

2a2. Reliability Testing **Testing attachment**

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.

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

- To test reliability, the developer used data obtained from 3,502 hospitals between 2013-2015.
- *Measure score reliability*
  - To demonstrate measure score reliability, the developer conducted a test/retest analysis to evaluate the measure's ability to generate consistent results with randomly selected subset of patients over time.
  - The developers calculated two metrics of agreement – the intraclass correlation coefficient (ICC) and the Spearman-Brown Prophecy Formula (S-B). The ICC is estimated from a random effects model producing risk adjusted rates. The S-B formal projects correlation as if the full sample is used and not spilt randomly.

Results of reliability testing

- The reliability testing results for the three-year period (CY2013-CY2015) produced an ICC of 0.570 (95% CI: 0.567, 0.572) and 0.482 (95% CI: 0.479, 0.485), for unadjusted and risk-adjusted values, respectively. The developer notes that this result may be interpreted as "fair" reliability.
- The mean S-B for the same period was 0.726 (95% CI: 0.724, 0.728) for unadjusted rates and 0.650 (95% CI: 0.648, 0.653) for risk-adjusted rates. The developer notes that both of these values are significantly higher than the 0.5 that indicates a large effect size with p-values < 0.001. When applied to each year individually, the S-B analysis exceeded 0.50 (p-values<0.001) in 2013 and 2014 but not 2015.

Questions for the Committee:

- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient reliability so that differences in performance can be identified?*

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.

Describe any updates to validity testing:

For this updated submission, the developer conducted additional validity testing by examining the measure score's correlation with other endorsed measures of readmissions.

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 conducted two new analyses to test the validity of the measure score. These analyses were:
  1. evaluating the sensitivity and specificity of the UB-04 inpatient admission type code. This analysis was previously conducted using a manual chart review.
  2. correlation between this measure and NQF #1789 CMS Hospital-Wide All-Cause Readmissions measure.

Validity testing results:

- The results of the two analysis are as follows:
  1. The previous data element validity testing generated a global sensitivity and specificity score of 0.879 and 0.896, respectively.
  2. The overall correlation between NQF #1789 and NQF #3188 was 0.2769 with a p-value of <0.001. This is a statistically significant positive correlation between the two measures.

Questions for the Committee:

- *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 following index admissions are excluded from the measure denominator:
  1) Less than 18 years of age (N=117, 0%);
  2) Patients who died during the index admission (N=200,855, 5.97%);
  3) Patients discharged AMA (N=12,612, 0.37%);
  4) Patients transferred to another acute care hospital during the index admission (N=78,692, 2.34%);
  5) Patients discharged with a planned readmission (N=3,970, 0.12%);
  6) Patients having missing or incomplete data (N=123, 0%) ; and,
  7) Patients not admitted to an inpatient bed (N=0, 0%).
- The following numerator exclusions are noted:
  - Readmissions for patients with progression of disease, defined as Primary Claim Diagnosis Code of metastatic disease (ICD-9-CM range: 196-198.89, 209.70 - 209.79; ICD-10-CM range: C77.0 – C79.9, C7B.0-C7B.8) (N=30,642, 4.18%)
    - Developer Rationale: A primary (or principal) diagnosis of metastatic disease serves as a proxy for

disease progression. Readmissions for conditions or symptoms associated with disease progression are not reflective of poor clinical care but, rather, advanced disease.
    - Readmissions for patients with planned admissions for treatment, defined as Primary Claim Diagnosis Code of chemotherapy or radiation encounter (ICD-9-CM range: V58.00-V58.12; ICD-10-CM range: Z51.00 – Z51.12). (N=19,028, 2.60%)
        - Developer Rationale: Readmissions are expected and planned for some patients who require additional cancer treatment in the inpatient setting. These readmissions reflects high-quality care that is focused on patient safety and are reliably distinguishable in claims data.
    - Adjust numerator to remove duplicate counts for multiple readmissions within the 30 day period (N=95,064, 12.98%)
- The developer provides frequency distributions and written rationale to justify exclusions. Information on performance results for patients excluded is not provided.

Questions for the Committee:

- *Are the exclusions consistent with the measure intent?*
- *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?*
- *Is there any concern that exclusions may create distortion of performance results across measured entities?*

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** ☐ **No**

Risk adjustment summary

- The statistical risk adjustment model includes 11 risk factors with 15 values.
- The developers use a logistic regression to estimate the probability of an unplanned readmission, based on the measure specifications and risk factors below:

Parameter-redo all numbers	Model Coefficients		Odds Ratio Estimates		
	Estimate	P-Value	Point Estimate	95% Wald Confidence Limits	
Intercept	-2.966	<.0001			
ICU Stay	0.055	<.0001	1.117	1.106	1.127
Male	0.046	<.0001	1.097	1.088	1.106
Dual-Eligible Status	0.069	<.0001	1.147	1.135	1.159
Surgical Admission	-0.226	<.0001	0.637	0.631	0.643
Multiple Comorbidities	0.123	<.0001	1.279	1.266	1.293
Solid Tumor *(excluding Metastatic Disease)*	-0.079	<.0001	0.854	0.847	0.861
Length of Stay Greater than 3 Days	0.149	<.0001	1.347	1.335	1.360
Age: < 65	*Reference Age*				
Age: 65-69	-0.075	<.0001	0.861	0.849	0.874
Age: 70-74	-0.068	<.0001	0.873	0.860	0.885
Age: 75-79	-0.078	<.0001	0.856	0.844	0.869
Age: 80-84	-0.101	<.0001	0.818	0.805	0.831
Age: 85+	-0.162	<.0001	0.723	0.712	0.735
Hospitalization in the Prior 60 Days	0.239	<.0001	1.612	1.597	1.627

Discharged to Home	-0.109	<.0001	0.804	0.797	0.811
Discharged to Hospice	-1.277	<.0001	0.078	0.075	0.080

Empirical Summary of SDS

- The developers noted that there was a conceptual and empirical rationale for adjustment based on dual-eligibility status. Dual-eligibility can serve as a proxy for low income status and other measures of SDS. Several studies were referenced that note that low SDS factors are a risk factor for later-state cancer diagnosis, delayed health care receipt, and higher utilization of hospital-based care.
- The patient-level observed 30-Day Unplanned Readmissions for Cancer Patients rate was 22.49%, compared with an 18.32% observed rate for all other patients.
- "Dual-Eligible Status" was associated with a Chi-Square of 5547.9628 ($p<0.001$).
- "Dual-Eligible Status" was included in the risk adjustment model.

Risk Model Discrimination and Calibration

- The developer provides a c-statistic and a Hosmer-Lemeshow statistic to assess risk adjustment model performance. The c-statistic measures how well the model discriminates between patients with and without the outcome, when compared with random assignment. A c-statistic of 0.5 suggests that the model has poor predictive power, while a c-statistic of 1.0 implies that the outcome is solely related to patient-level factors. The c-statistic provided by the developer 0.6607 (95% CI: 0.6597, 0.6618), indicating fair discrimination for the development and validation models.
- The H-L Goodness-of-Fit test yielded a significant value ($p<0.001$), which indicates potential fit issues.
- The developer notes that this is not uncommon with models that are overpowered due to large datasets, as is the case here. A significant value for the H-L test suggests that we reject the assumption of perfect fit between the models. However, with large datasets, the H-L statistic can magnify relatively small differences between observed and expected rates and imply a statistically significant degree of miscalibration.
- The developer notes that the risk decile plots demonstrate that the model performs adequately, with similar observed and predicted values in each decile.

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?*
- *Are all of the risk adjustment variables present at the start of care? If not, describe the rationale provided.*
- *Do you agree with the developer's rationale that there is a conceptual basis for adjusting this measure for SDS factors?*
- *Do you agree with the developer's decision, based on their analysis, to 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)*:

- To demonstrate the measure's ability to identify meaningful differences, the developer compared hospital unadjusted and adjusted rates compared to the mean national performance rates.
- Half of the hospitals fell within the interquartile range of 12.50% to 20.80%
- The developers note that in their analysis of total and for CY2015 individually, they observed that over half of all index claims had performance of "no better or worse than the national average" – demonstrating that there are opportunities for improvement by providers.
-

Question for the Committee:

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

2b6. Comparability of data sources/methods:

N/A

2b7. Missing Data

- The developer states that all required data are readily available and retrievable. Missing data does not appear to be an issue for this measure.

Guidance from the Validity Algorithm
Precise specifications (Box 1) → Empirical testing conducted with measure as specified (Box 2) → Score-level testing conducted (Box 4) → Validity testing for each measured entity (Box 6) Method of testing appropriate (Box 7) → moderate certainty that the scores are valid

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:

** Used 3502 hospitals between 2013-2015. Did test-retest, ICC and S-B. producing unadjusted and risk-adjusted values. Moderate reliability. This measure can be consistently implemented.

** Specifications clear: yes; Elements well defined: yes; logic clear: yes;

** Reliability specs acceptable

** Specs appear reasonable. Question: has there been reliability testing re: use of metastatic disease as a primary diagnosis as a proxy for disease progression? Might posit that some percentage of patients have metastatic disease at d/c from the index hospitalization and exclusion might obscure transitional care issues in these patients.

** As data specs are electronically abstractable from claims, likely no issues with implementation.

** Clearly defined.

Unplanned readmission not by coding, rather exclude urgent/emergent admission types. These may not accurately describe unplanned readmissions and does not follow current methodology used by similar measures

** Recognizing that some of the risk factors identified were removed from consideration because they were not well-defined in the claims data, many of these are significant contributors to patients' ability to follow through on their post hospitalization care. For example, history of substance abuse and psychological services are major contributors to readmissions and by removing these, hospitals that serve high numbers of these patients are placed at a significant disadvantage in accurately measuring risk.

** No issues

** Included readmission for emergency or urgent. Unclear if included observation

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:

** Empirical testing done using 2 new analyses (as compared to previous submission with measure 2884?). Sensitivity and specificity of UB-04 IP admissions type code and correlation of measure with CMS all cause HW readmissions (NQF 1789). Global sensitivity and specificity high and found a significant correlation with NQF 1789. No specifications inconsistent with the evidence.

** Consistent with evidence: yes

** Validity specs consistent

** No identified issues

**Moderate

** No issues noted

** No issues

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:

** Appropriate method and large number of hospitals. Moderate reliability, appropriate method and generalizable.

** Sample adequate: yes; Differences in performance can be identified: yes; Reliability rating: pass

** Testing was at facility level for 4QCY2012 to 1QCY2016 encompassing 4,974 short term acute care hospitals. There was a minimum threshold of 50 index admissions needed for inclusion. On page 29, there appear to have been few PPS exempt hospitals included in testing group.

** Reliability testing acceptable- moderate

** Sample size for some hospitals is small--probably useful to exclude hospitals without the minimum number of index admits.
Split half testing, ICC scores were in the fair range.

** No issues noted.

** The developer states that data from 3,502 hospitals was used in their reliability testing. No mention is made about total number of patients/records, however.

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:

** Sufficient validity for conclusions regarding quality of care and readmissions.

** Sufficient validity: yes; sensitivity and specificity: good; positive correlation with similar measures: good

** Validity testing acceptable- moderate

** Interesting use of performance on NQF 1789 as a measure of validity for the identification of type of admission: sens/spec in good ranges.

** Adequate

** No issues.

** OK

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:

** No

** Exclusions consistent with intent: yes; patients are appropriately excluded: yes; Exclusions needed: yes; Potential distortion from exclusions: no; Appropriate risk adjustment strategy: yes; Variable adequately described: yes; Variables present: yes; Agree with developer's rationale for SDS adjustment: yes; Agree with decision to include SDS: yes; Meaningful differences in quality: yes; Missing data problem: no

** There were no comparable measures that could be used for comparison due to gap.

** Acceptable

** 2b3: understanding the developers' point re: metastatic dz as a proxy for disease progression, what percentage of index hospitalizations included diagnoses for metastatic dz--is it always a measure of disease progression vs disease acuity at dx? Exclusions for death, missing data, AMA, transfer seemed reasonable; frequency of denominator exclusions seemed low.

2b4: developers allude to a possible conceptual relationship between low SES and readmissions; used dual eligibility as a proxy for SDS factors. Also, race was removed as a variable 2/2 potential to mask disparities in care, also developers did not think they could articulate a causal relationship between race and readmissions.

2b5: narrow IQR, majority of performance in test sample was around national average. Histograms seemed to indicate

some outliers and potential for quality improvement.

** Missing data is listed as potential exclusion

** No

**No

** Narrow focus

** Fair reliability based on ICC

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.

- This measure is calculated using administrative claims data from established data fields. Thus, the measure's required data elements are routinely generated 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:
Are the required data elements routinely generated and used during care delivery?

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

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:

** This is a highly feasible measure given the source of data.

** Feasible: yes

** I think data can be collected since it is Medicare claims data.

** High

** Administrative claims data: no issues with feasibility

** Data routinely available. increase in coding around "planned" readmission

** The majority of the data elements are readily available and documented in the electronic medical record. Should not be difficult to replicate to use on an operational level.

** None

** OK

** 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**

Quality Improvement

- The measure is publically reported by Vizient, Inc. with external benchmarking to multiple organizations.
- The developer notes that the measure is also used in quality improvement applications at the City of Hope Comprehensive Care Center, University of Miami Sylvester Comprehensive Cancer Care, Seattle Cancer Care Alliance

Accountability Applications

- The measure is used in the Annual Hospital Ratings for Colon and Lunch Cancer Surgery.
- The measure is used in an ACO payment program at Moffitt Cancer Center with Florida Blue.

Improvement results N/A

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

Potential harms N/A

Vetting of the measure [vetting] N/A

Feedback:

- The All-Cause Admission and Readmissions Standing Committee reviewed #2884 (the measure from which #3188 is adapted) during the 2015-2016 evaluation cycle. #2884 was not recommended for endorsement due to limitations related to care setting and measure testing. #3188 addresses the Standing Committee's recommendation by broadening the measure to capture cancer patient readmissions to any short-term acute care PPS hospital and by conducting additional testing using Medicare claims data.
- This measure was included in CMS' 2014 Measures Under Consideration (MUC) list and received conditional support from the Measure Applications Partnership (MAP) Hospital Work Group, pending NQF endorsement. The developer expects the measure to be included in future rule-making; potentially as early as the FY 2018 Hospital Inpatient PPS Proposed Rule.

***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**

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:

** At this time being used 3 institutions for QI and 2 accountability applications.

** Currently in use: yes; Unintended consequences: Hospitals that reduce cancer readmissions may increase post discharge mortality; Measures of readmission and mortality should be monitored and reported in tandem; Validity testing of readmission measure should include mortality as a variable.

** it appears the Hospital-wide All-Cause Unplanned Readmission measure excludes non-surgical cancer admissions and PCHs.

I would like to better understand the size of the exclusion group since patients readmitted for chemotherapy, radiation therapy treatment or with disease progression are excluded. It is important to note that most common reason for readmissions appear to be infections, fever and gastro-intestinal complications."

** Moderate

** Similar measures already being used internally by a variety of cancer hospitals and at least 1 ACO.

** Measure is improved as compared to previously submitted by including readmissions beyond only cancer hospitals

** As with other readmission measures, SDS factors are not appropriately measured leaving inner city hospitals at a disadvantage when publicly reporting readmission rates. Although dual-eligibility status is a proxy, there are many factors that contribute and are not taken into account using a proxy. Availability of community services, such as a robust public transportation system as one example, significantly impacts patients' ability to obtain follow up care including medications. This creates a high use of the 911 system for those patients needing care. Although significant effort may be taken to safely discharge a patient out of the hospital, the lack of available resources often leads to non-adherence in the treatment plan.

** The measure is at least as feasible as existing readmissions measures. Assuming the committee agrees with the rationale to report readmissions data for cancer patients independent from broader, all-cause readmissions measures, feasibility isn't a concern of mine.

** OK

** Publically reported by Vizient, Inc.
Used by 3 comprehensive cancer centers for quality improvement applications

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 measure has not been in use or broadly vetted by those being measured or other users.

Pre-meeting public and member comments

-

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** Click here to enter NQF number

Measure Title: 30-Day Unplanned Readmissions for Cancer Patients

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: 1/13/2017

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 Unplanned Readmissions for Cancer Patients

☐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

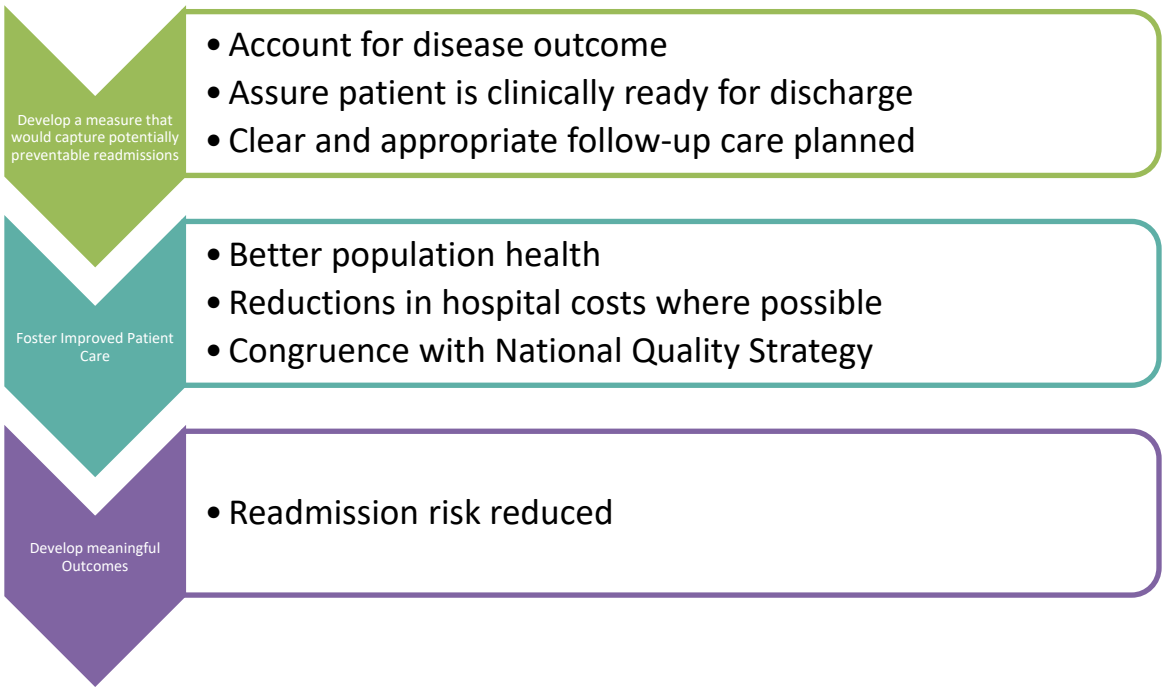
☐ Process: Click here to name what is being measured

☐ 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.



This measure was developed to yield risk-adjusted, hospital-level rates of unplanned readmissions that:

1) Are valid and reliable for cancer care;
2) Address cancer measurement gaps in existing readmissions measures;

3) Are capable of differentiating quality of care;
4) Are useful for quality improvement; and,
5) May be used in public reporting programs to inform patients, payers, and policymakers regarding the quality of hospital-based cancer care.

Using a broad Medicare claims set, patients with a Type of Admission/Visit of "emergency" or "urgent" within 30 days of an index admission are considered unplanned readmissions in the measure. The measure excludes readmissions for patients readmitted for chemotherapy or radiation therapy treatment or with disease progression.

By providing an accurate and comprehensive assessment of unplanned readmissions within 30 days of discharge, hospitals can better identify and address preventable readmissions. Through routine use, this measure can be used to improve patient outcomes and quality of care. The measure is intended to identify institutions that are performing better or worse than expected and to support improved care delivery and quality of life for this complex patient population.

While measure testing has focused on producing a measure that can be applied to PPS-Exempt Cancer Hospitals (PCH), we believe that the measure has broad applicability to cancer patients treated in any short-term acute care hospital. Accordingly, this measure could be adopted for the PPS-Exempt Cancer Hospitals Quality Reporting Program (PCHQR) and other public reporting programs for purposes of accountability and to support performance improvement.

****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).

Cancer is the second leading cause of death in the United States, with nearly 600,000 cancer-related deaths expected this year.[1] It is now the leading cause of death among adults aged 40 to 79 years as well and in 21 states.[2] It is estimated roughly 1.7 million Americans will be diagnosed with cancer in 2016, and nearly 14.5 million Americans with a history of cancer were alive in 2014. Cancer disproportionately affects older Americans, with 86% of all cancers diagnosed in people 50 years of age and older.[1] Oncology care contributes greatly to Medicare spending and accounted for an estimated $125 billion in healthcare spending in 2010. This figure is projected to rise to between $173 billion and $207 billion by 2020.[3] Given the current and projected increases in cancer prevalence and costs of care, it is essential that healthcare providers look for opportunities to lower the costs of cancer care.

Reducing readmissions after hospital discharge has been proposed as an effective means of lowering healthcare costs and improving the outcomes of care. Research suggests that between 9% and 48% of all hospital readmissions are preventable, owing to inadequate treatment during the patient's original (index) admission or after discharge.[4] Jencks, et al. estimated that unplanned readmissions cost the Medicare program $17.4 billion in 2004.[5]

Unnecessary hospital readmissions negatively impact cancer patients by compromising their quality of life, by placing them at risk for health-acquired infections, and by increasing the costs of their care. Furthermore,

unplanned readmissions during treatment can delay treatment completion and, potentially, worsen patient prognosis.

Preventing these readmissions improves the quality of care for cancer patients. Numerous studies have examined all-cause readmissions and readmissions for specific conditions, such as orthopedic surgery. Existing studies in cancer have largely focused on post-operative readmissions, reporting readmission rates between 6.5% and 25%. Patient factors, including age, comorbidities, cancer stage, and socioeconomic status, were identified as risk factors in these patients. Surgical complications, surgery duration, and hospital length of stay also increased readmission risk in these studies. Finally, hospital factors (e.g., hospital size) and practice patterns, such as inadequate discharge planning, comorbidity management, and follow-up care, were associated with preventable readmissions.[6-17] Moya, et al. observed a 20% readmission rate in hematopoietic cell transplantation (HCT) recipients along with an extended length of stay during the readmission (25 ± 21 days). Infections (some associated with the graft), graft failure, coagulation disorders, and a second neoplasm were the most frequent causes of readmission.[18] Bejanyan, et al. examined readmissions in patients with myeloablative allogeneic HCT and observed a 39% readmission rate in these patients. Infections, fever, gastrointestinal complications, and graft-versus-host disease (GVHD) were the most frequent reasons for readmission.[19] Less is known about other readmissions in medical cancer admissions, though Ji, et al. noted that surgical patients were most often readmitted for surgical complications while medical patients were typically readmitted for the same condition treated during the index admission.[6] Together, these studies suggest that certain readmissions in cancer patients are preventable and should be routinely measured for purposes of quality improvement and accountability.

All-cause and disease-specific unplanned readmissions rates have been adopted by the Centers for Medicare & Medicaid Services (CMS) as key indicators of inpatient quality care. Additionally, Medicare began reducing payments to hospitals with excess readmissions in October 2012, as mandated in the Patient Protection and Affordable Care Act of 2010. Benbassat, et al. concluded that global readmission rates are not useful indicators of healthcare quality and, instead, recommended measuring readmissions at the condition level.[4] Readmission rates have been developed for pneumonia, acute myocardial infarction, and heart failure. However, cancer has lagged behind these conditions in the development of validated readmission rates. In 2012, the Comprehensive Cancer Center Consortium for Quality Improvement, or C4QI (a group of eighteen academic medical centers that collaborate to measure and improve the quality of cancer in their centers), began development of a cancer-specific unplanned readmissions measure: *30-Day Unplanned Readmissions for Cancer Patients*. The Alliance of Dedicated Cancer Centers, or ADCC (an organization of eleven comprehensive cancer centers that are reimbursed differently by Medicare), identified this ongoing work as a potential accountability measure for the PCHQR. Both groups recognize the importance of measuring unplanned readmissions as an indicator of the quality of hospital-based oncology care and have designed the *30-Day Unplanned Readmissions for Cancer Patients* measure accordingly.[5,6] This measure is intended to reflect the unique clinical aspects of oncology patients and to yield readmission rates that more accurately reflect the quality of cancer care delivery, when compared with broader readmissions measures. Likewise, this measure addresses cancer measurement gaps in existing readmissions measures, such as the *Hospital-Wide All-Cause Unplanned Readmission Measure (HWR)*, stewarded by CMS. The *30-Day Unplanned Readmissions for Cancer Patients* measure can be used by individual hospitals to inform local quality improvement efforts. Through adoption in public reporting programs (e.g., PCHQR), it can increase transparency around the quality of care delivered to patients with cancer.

REFERENCES

1. American Cancer Society. Cancer facts and figures 2016. 2016. Available at: http://www.cancer.org/acs/groups/content/@research/documents/document/acspc-047079.pdf.

2. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2016. *CA Cancer J Clin.* 2016;66(1):7-30.

3. Mariotto AB, Yabroff KR, Shao Y, Feuer EJ, Brown ML. Projections of the cost of cancer care in the United States: 2010-2020. *J Natl Cancer Inst.* 2011;103(2):117-128.

4. Benbassat J, Taragin M. Hospital readmissions as a measure of quality of health care: advantages and limitations. *Arch Intern Med.* 2000;160(8):1074-1081.

5. Jencks SF, Williams MV, Coleman EA. Rehospitalizations among patients in the Medicare fee-for-service program. *N Engl J Med.* 2009;360(14):1418-1428.

6. Ji H, Abushomar H, Chen XK, Qian C, Gerson D. All-cause readmission to acute care for cancer patients. *Healthc Q.* 2012;15(3):14-16.

7. Rochefort MM, Tomlinson JS. Unexpected readmissions after major cancer surgery: an evaluation of readmissions as a quality-of-care indicator. *Surg Oncol Clin N Am.* 2012;21(3):397-405, viii.

8. Manzano JG, Luo R, Elting LS, George M, Suarez-Almazor ME. Patterns and predictors of unplanned hospitalization in a population-based cohort of elderly patients with GI cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology.* 2014;32(31):3527-3533.

9. Dickinson H, Carico C, Nuno M, et al. Unplanned readmissions and survival following brain tumor surgery. *J Neurosurg.* 2015;122(1):61-68.

10. Fernandez FG, Khullar O, Force SD, et al. Hospital readmission is associated with poor survival after esophagectomy for esophageal cancer. *Ann Thorac Surg.* 2015;99(1):292-297.

11. Manzano JG, Gadiraju S, Hiremath A, Lin HY, Farroni J, Halm J. Unplanned 30-Day Readmissions in a General Internal Medicine Hospitalist Service at a Comprehensive Cancer Center. *J Oncol Pract.* 2015;11(5):410-415.

12. Saunders ND, Nichols SD, Antiporda MA, et al. Examination of unplanned 30-day readmissions to a comprehensive cancer hospital. *J Oncol Pract.* 2015;11(2):e177-181.

13. Shah SP, Xu T, Hooker CM, et al. Why are patients being readmitted after surgery for esophageal cancer? *J Thorac Cardiovasc Surg.* 2015;149(5):1384-1389; discussion 1389-1391.

14. Valero-Elizondo J, Kim Y, Prescott JD, et al. Incidence and Risk Factors Associated with Readmission After Surgical Treatment for Adrenocortical Carcinoma. *J Gastrointest Surg.* 2015;19(12):2154-2161.

15. Uppal S, Penn C, Del Carmen MG, Rauh-Hain JA, Reynolds RK, Rice LW. Readmissions after major gynecologic oncology surgery. *Gynecol Oncol.* 2016;141(2):287-292.

16. Wilbur MB, Mannschreck DB, Angarita AM, et al. Unplanned 30-day hospital readmission as a quality measure in gynecologic oncology. *Gynecol Oncol.* 2016;143(3):604-610.

17. Nakayama JM, Ou JP, Friedman C, Smolkin ME, Duska LR. The Risk Factors of Readmission in Postoperative Gynecologic Oncology Patients at a Single Institution. *Int J Gynecol Cancer.* 2015;25(9):1697-1703.

18. Moya R, Espigado I, Parody R, Carmona M, Marquez F, De Blas JM. Evaluation of readmissions in hematopoietic stem cell transplant recipients. *Transplant Proc.* 2006;38(8):2591-2592.

19. Bejanyan N, Bolwell BJ, Lazaryan A, et al. Risk factors for 30-day hospital readmission following myeloablative allogeneic hematopoietic cell transplantation (allo-HCT). *Biol Blood Marrow Transplant.* 2012;18(6):874-880.

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 sub criteria to pass this criterion and be evaluated against the remaining criteria.***
1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form 2017_01_13_UnplannedReadm_Cancer_NQF_evidence_attachment_Final.docx **1a.1 For Maintenance of Endorsement: Is there new evidence about the measure since the last update/submission?** Please update any changes in the evidence attachment in red. Do not remove any existing information. If there have been any changes to evidence, the Committee will consider the new evidence. If there is no new evidence, no updating of the evidence information is needed.
1b. Performance Gap Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating: • considerable variation, or overall less-than-optimal performance, in the quality of care across providers; and/or

- Disparities in care across population groups.

1b.1. Briefly explain the rationale for this measure *(e.g., how the measure will improve the quality of care, the benefits or improvements in quality envisioned by use of this measure)*
IF a 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.)
IF a COMPOSITE (e.g., combination of component measure scores, all-or-none, any-or-none), SKIP this question and provide rationale for composite in question 1c.3 on the composite tab.
For many cancer patients, readmission following hospitalization may be preventable and should be addressed to potentially lower costs and improve patient outcomes. The Alliance of Dedicated Cancer Centers, or ADCC (an organization of the eleven National Cancer Institute-designated comprehensive cancer centers that are exempt from the Prospective Payment System), recognizes the need for oncology-specific efficiency measures, including unplanned readmissions because planned readmissions are often used in clinical pathways for cancer patients. In 2014, the ADCC identified the 30-Day Unplanned Readmissions for Cancer Patients measure as a potential accountability measure for the PPS-Exempt Cancer Hospitals Quality Reporting Program (PCHQR). The measure was initially developed by the Comprehensive Cancer Centers for Quality Improvement (C4QI), a group of twenty-one academic medical centers that collaborate to measure and improve the quality of cancer care in their institutions. C4QI's 21 members (11 ADCC hospitals/PCHs and 10 other academic medical centers, or AMC) have utilized this claims-based, cancer-specific unplanned readmissions measure since 2012. It is designed to reflect the unique clinical aspects of oncology and to provide a more comprehensive measurement of unplanned readmissions in cancer patients, when compared with existing measures (e.g., the HWR measure). It considers patients with an admission type of "emergency" or "urgent" within 30 days of an index admission as an unplanned readmission. It excludes readmissions for patients readmitted for chemotherapy or radiation therapy treatment or with disease progression. Using this measure, hospitals can better identify and address preventable readmissions for cancer patients.

An earlier version of this measure (NQF #2884) was reviewed by the NQF All-Cause Admissions and Readmissions Project 2015-2017 Technical Expert Panel (TEP) in June 2016. Following the recommendation of the TEP, the ADCC broadened the measure to capture readmissions of cancer patients from and to any short-term acute care PPS hospital and pursued additional testing of the measure using Medicare claims data (i.e., the Standard Analytical Files). This expansion produced unplanned readmissions rates of patients discharged from PCHs and readmitted to any short-term acute care hospital (defined as PCHs, short-term acute care Prospective Payment System, or PPS, hospitals, and Critical Access Hospitals, or CAH). Additionally, it provided comparative rates of unplanned readmissions of cancer patients for non-PCH short–term acute care hospitals (i.e., short-term acute care PPS hospitals and CAHs).

1b.2. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis. *(This is required for maintenance of endorsement. Include mean, std dev, min, max, interquartile range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the sub-criterion on improvement (4b) under Usability and Use.*
30-Day Unplanned Readmissions for Cancer Patients
All Short-Term Acute Care Hospitals
CY2013-2015 Summary Statistics-Unadjusted Rates

	2013-15	2013	2014	2015
Number of Hospitals		4,974	4,736	4,722	4,688
Number of Admissions (Denominator)		3,067,675	1,037,916	1,016,301	1,013,458
Number of Unplanned Readmissions (Numerator)		587,915	198,039	194,993	194,883
30-Day Unplanned Readmission Rate		19.16%	19.08%	19.19%	19.23%
Mean (Standard Deviation)		16.54% (8.24%)	16.53% (10.36%)	16.47% (10.71%)	16.64% (11.01%)
Range (Min-Max)		0.00%-100.00%	0.00%-100.00%	0.00%-100.00%	0.00%-100.00%
Quartile Range	8.30%	10.32%	10.32%	10.53%	
Minimum	0.00%	0.00%	0.00%	0.00%	
25th percentile	12.50%	11.11%	11.11%	11.11%	
50th percentile	17.32%	17.20%	17.23%	17.35%	
75th percentile	20.80%	21.43%	21.43%	21.64%	
Maximum	100.00%	100.00%	100.00%	100.00%	

Table 1: Summary-level statistics for the 30-Day Unplanned Readmissions for Cancer Patients measure—shows unadjusted results of the 30-Day Unplanned Readmissions for Cancer Patients measure, when applied to 1Q CY2013-4Q CY2015 index admissions for short-term acute care hospitals (i.e., PCHs, short-term acute care PPS hospitals, and CAHs). Data source: Analysis of Medicare SAF (4Q2012-1Q2016), based on data provided by Watson Policy Analysis, 01/13/2017.

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.

1b.4. Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability. *(This is required for maintenance of endorsement. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included.) For measures that show high levels of performance, i.e., "topped out", disparities data may demonstrate an opportunity for improvement/gap in care for certain sub-populations. This information also will be used to address the sub-criterion on improvement (4b) under Usability and Use.*

Measure testing produced the following descriptive statistics for patient-level demographic variables, which were evaluated in our risk adjustment model:

For the denominator:

Sex

Value	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Unknown	5,560	0.18%	5,560	0.18%
Male	1,616,259	52.69%	1,621,819	52.87%
Female	1,445,856	47.13%	3,067,675	100.00%

Table 2: "Sex" variable distribution for the 30-Day Unplanned Readmissions for Cancer Patients measure—includes the distribution of the "Sex" variable for the denominator population when the 30-Day Unplanned Readmissions for Cancer Patients measure is applied to 1Q CY2013-4Q CY2015 index admissions for all 4,974 short-term acute care hospital (defined as PCHs, short-term acute care PPS hospitals, and CAHs). Data source: Analysis of Medicare SAF (4Q2012-1Q2016), based on data provided by Watson Policy Analysis, 01/13/2017.

Age at Beginning of Reference Year

Value	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Unknown	5,560	0.18%	5,560	0.18%
Under 65	409,844	13.36%	415,404	13.54%
65-69	618,508	20.16%	1,033,912	33.70%
70-74	606,147	19.76%	1,640,059	53.46%
75-79	529,837	17.27%	2,169,896	70.73%
80-84	424,681	13.84%	2,594,577	84.58%
85+	473,098	15.42%	3,067,675	100.00%

Table 3: "Age " variable distribution for the 30-Day Unplanned Readmissions for Cancer Patients measure—includes the distribution of the "Age" variable for the denominator population when the 30-Day Unplanned Readmissions for Cancer Patients measure is applied to 1Q CY2013-4Q CY2015 index admissions for all 4,974 short-term acute care hospital (defined as PCHs, short-term acute care PPS hospitals, and CAHs). The "Age" variable is populated by adding one year to the "Age" field in the Medicare SAF (1Q2013-4Q2015), which is reported as the beneficiary's age at the end of the prior year. Data source: Analysis of Medicare SAF (4Q2012-1Q2016), based on data provided by Watson Policy Analysis, 01/13/2017.

Beneficiary Race Code

Value	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Unknown	28,494	0.93%	28,494	0.93%
White	2,535,852	82.66%	2,564,346	83.59%
Black	354,140	11.54%	2,918,486	95.14%
Other	39,428	1.29%	2,957,914	96.42%
Asian	42,990	1.40%	3,000,904	97.82%
Hispanic	52,158	1.70%	3,053,062	99.52%

North American Native	14,613	0.48%	3,067,675	100.00%

Table 4: "Race" variable distribution for the 30-Day Unplanned Readmissions for Cancer Patients measure—includes the distribution of the "Race" (or "Beneficiary Race Code") variable for the denominator population when the 30-Day Unplanned Readmissions for Cancer Patients measure is applied to 1Q CY2013-4Q CY2015 index admissions for all 4,974 short-term acute care hospital (defined as PCHs, short-term acute care PPS hospitals, and CAHs). Data source: Analysis of Medicare SAF (4Q2012-1Q2016), based on data provided by Watson Policy Analysis, 01/13/2017.

Dual-Eligible Status

Value	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Never dual eligible	2,448,890	79.83%	2,448,890	79.83%
Dual eligible at some point	618,785	20.17%	3,067,675	100.00%

Table 5: "Dual-Eligible Status" variable distribution for the 30-Day Unplanned Readmissions for Cancer Patients measure—includes the distribution of dual-eligible Medicare beneficiaries for the denominator population when the 30-Day Unplanned Readmissions for Cancer Patients measure is applied to 1Q CY2013-4Q CY2015 index admissions for all 4,974 short-term acute care hospital (defined as PCHs, short-term acute care PPS hospitals, and CAHs). The "Dual-Eligible Status" variable is used as a proxy for socioeconomic status and is populated by analyzing the Buyin field in the Medicare SAF (1Q2013-4Q2015). Patients with any claims with a value of "A", "B", or "C" in the Buyin field in the 1Q2013-4Q2015 data set are coded as "Dual-Eligible" in this variable. All other patients are coded as "Never Dual-Eligible" in this variable. Data source: Analysis of Medicare SAF (4Q2012-1Q2016), based on data provided by Watson Policy Analysis, 01/13/2017.

1b.5. **If no or limited data on disparities from the measure as specified is reported in 1b.4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations. Not necessary if performance data provided in 1b.4**

2. Reliability and Validity—Scientific Acceptability of Measure Properties

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

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

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

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

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

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.)*

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:** 2017_01_13_UnplannedReadm_Cancer_DataDictv1.0.xls

S.3.1. **For maintenance of endorsement:** Are there changes to the specifications since the last updates/submission. If yes, update the specifications for S1-2 and S4-22 and explain reasons for the changes in S3.2.

S.3.2. **For maintenance of endorsement,** please briefly describe any important changes to the measure specifications since last measure update and explain the reasons.

S.4. **Numerator Statement** *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome) DO NOT include the rationale for the measure.*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).
This outcome measure demonstrates the rate at which adult cancer patients have unplanned readmissions within 30 days of discharge from an eligible index admission. The numerator includes all eligible unplanned readmissions to any short-term acute care hospital—defined as admission to a PPS-Exempt Cancer Hospital (PCH), a short-term acute care Prospective Payment (PPS) hospital, or Critical Access Hospital (CAH)—within 30 days of the discharge date from an index admission that is included in the measure denominator. Readmissions with an admission type (UB-04 Uniform Bill Locator 14) of "emergency = 1" or "urgent = 2" are considered unplanned readmissions within this measure. Readmissions for patients with progression of disease (using a principal diagnosis of metastatic disease as a proxy) and for patients with planned admissions for treatment (defined as a principal diagnosis of chemotherapy or radiation therapy) are excluded from the measure numerator.

S.5. **Numerator Details** *(All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
IF an OUTCOME MEASURE, describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).
The numerator includes readmissions of the following patients with an eligible index admission in the measure denominator:
1) Readmitted to a short-term acute care hospital (PCHs, short-term acute care PPS hospitals, and CAHs) within 30 days of the discharge date of an index admission; and,
2) Readmitted with a Claim Inpatient Admission Type Code of "Emergency" or "Urgent" ("1" or "2").

The following readmissions are excluded from the measure numerator:
1) Primary Claim Diagnosis Code of metastatic disease (ICD-9-CM range: 196-198.89, 209.70-209.79; ICD-10-CM range: C77.0 – C79.9, C7B.0-C7B.8).
Rationale: A primary (or principal) diagnosis of metastatic disease serves as a proxy for disease progression. Readmissions for conditions or symptoms associated with disease progression are not reflective of poor clinical care but, rather, advanced disease.

2) Patients with a Primary Claim Diagnosis Code of chemotherapy or radiation encounter (ICD-9-CM range: V58.00-V58.12; ICD-10-CM range: Z51.00 – Z51.12) as these are considered planned admissions.
Rationale: Readmissions are expected and planned for some patients who require additional cancer treatment in the inpatient setting. These readmissions reflects high-quality care that is focused on patient safety and are reliably distinguishable in claims data.

Of note, if a patient has more than one unplanned admission within 30 days of discharge from the index admission, each readmission is only counted once in the numerator.

S.6. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
The denominator includes inpatient admissions for all adult Fee-for-Service Medicare beneficiaries where the patient is discharged from a short-term acute care hospital (PCH, short-term acute care PPS hospital, or CAH) with a principal or secondary diagnosis (i.e., not admitting diagnosis) of malignant cancer within the defined measurement period.

S.7. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b.)*
IF an OUTCOME MEASURE, describe how the target population is identified. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).
The denominator includes index admissions at acute care hospitals (PCHs, short-term acute care PPS hospitals, and CAHs) for patients with a discharge date during the measurement period that meet the following criterion:
1) Primary Claim Diagnosis Code or Claim Diagnosis Code I-XXV of malignant cancer (ICD-9-CM range: 140.00-209.36, 209.70-209.79, 511.81, 789.51; ICD-10-CM range: C00 – C96.9, J91.0, R18.0).

Of note, a readmission that meets the denominator criteria is included as an index admission within this measure if it meets all other eligibility criteria.

S.8. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
The following index admissions are excluded from the measure denominator:
1) Less than 18 years of age;
2) Patients who died during the index admission;
3) Patients discharged AMA;
4) Patients transferred to another acute care hospital during the index admission;
5) Patients discharged with a planned readmission;
6) Patients having missing or incomplete data; and,
7) Patients not admitted to an inpatient bed.

S.9. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b.)*
The following index admissions are excluded from the measure denominator:
1) Age less than 18 years of age (based on the beneficiary's age at the end of the prior year).
Rationale: Pediatric patients represent a very small and distinct Medicare population with different characteristics and outcomes.

2) Patient Discharge Status Code indicating "Expired" (20).
Rationale: Patients that die during the index admission cannot be readmitted.

3) Patient Discharge Status Code indicating "Left Against Medical Advice" (07).
Rationale: The hospital had limited opportunity to ensure the patient was prepared for discharge and had appropriate follow-up care.

4) Patient Discharge Status Code indicating transfer to an acute care facility (02, 05, 09, 30, 43, 66, 69).
Rationale: Responsibility for any unplanned readmissions is assigned to the final discharging hospital. Intermediate index admissions within a single episode of care are ineligible for inclusion.

5) Patient Discharge Status Code indicating discharge with a planned readmission (81-95).
Rationale: The patient was discharged with a planned readmission, which is ineligible for the measure numerator.

6) Patient Discharge Status Code indicating "Unknown Value" (0, 40-42) or Organization NPI Number = "".
Rationale: Admissions without a valid discharge status cannot be evaluated for measure exclusions. Admissions with a discharge status reserved for hospice claims only are not admissions for acute care or to acute care hospitals. Claims without an Organizational NPI Number cannot be evaluated for inclusion in the measure.

7) NCH Claim Type Code indicating a claim record type is not an "Inpatient Claim" (all values except 60).
Rationale: These admissions are not for acute care or to acute care hospitals.

S.10. **Stratification Information** *(Provide all information required to stratify the measure results, if necessary, including the stratification variables, definitions, specific data collection items/responses, code/value sets, and the risk-model covariates and coefficients for the clinically-adjusted version of the measure when appropriate – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format with at S.2b.)*

Measure is not stratified.

S.11. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in measure testing attachment)
Statistical risk model
If other:

S.12. **Type of score:**
Rate/proportion
If other:

S.13. **Interpretation of Score** *(Classifies interpretation of score according to whether better quality is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score)*
Better quality = Lower score

S.14. **Calculation Algorithm/Measure Logic** (*Diagram or describe the calculation of the measure score as an ordered sequence of steps including identifying the target population; exclusions; cases meeting the target process, condition, event, or outcome; time period for data, aggregating data; risk adjustment; etc.)*
Please refer to the measure flow logic in the data dictionary.

S.15. **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.
This outcome measure is based on the full population of eligible patients; sampling is not applied.

S.16. **Survey/Patient-reported data** *(If measure is based on a survey or instrument, provide instructions for data collection and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
N/A

S.17. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.18.
Claims (Only)

S.18. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument (e.g. name of database, clinical registry, collection instrument, etc., and describe how data is collected.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
The Medicare 100% Standard Analytic File (SAF) covering CY2013 through CY2016Q1 was used for testing purposes. This contains 100% of the claims for the Fee-for-Service population. The specific files used were the Inpatient file containing information on inpatient claims and the Denominator file containing information on the enrollment and demographics. As these data are released in separate files, the data files were combined by a statistician at Watson Policy Analysis for purposes of measure testing.

S.19. **Data Source or Collection Instrument** *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No data collection instrument provided

S.20. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Facility

S.21. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Hospital : Acute Care Facility
If other:

S.22. **COMPOSITE Performance Measure** - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*
N/A

2. Validity – See attached Measure Testing Submission Form
2017_01_13_UnplannedReadm_Cancer_NQF_testing_attachment_Final.docx

2.1 For maintenance of endorsement
Reliability testing: If testing of reliability of the measure score was not presented in prior submission(s), has reliability testing of the measure score been conducted? If yes, please provide results in the Testing attachment. (Do not remove prior testing information – include date of new information in red.)

2.2 For maintenance of endorsement
Has additional empirical validity testing of the measure score been conducted? If yes, please provide results in the Testing attachment. (Do not remove prior testing information – include date of new information in red.)

2.3 For maintenance of endorsement
Risk adjustment: For outcome, resource use, cost, and some process measures, risk-adjustment that includes SDS factors is no longer prohibited during the SDS Trial Period (2015-2016). Please update sections 1.8, 2a2, 2b2, 2b4, and 2b6 in the Testing attachment and S.14 and S.15 in the online submission form in accordance with the requirements for the SDS Trial Period. NOTE: These sections must be updated even if SDS factors are not included in the risk-adjustment strategy. If yes, and your testing attachment does not have the additional questions for the SDS Trial please add these questions to your testing attachment:

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).

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)

What were the statistical results of the analyses used to select risk factors?

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)

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

Measure Number (*if previously endorsed*)**:** Click here to enter NQF number
Measure Title: 30-Day Unplanned Readmissions for Cancer Patients
Date of Submission: 1/13/2017
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 <u>all</u> measures, sections 1, 2a2, 2b2, 2b3, and 2b5 must be completed.**
- **For <u>outcome and resource use</u> measures**, section **2b4** also must be completed.
- If specified for **<u>multiple data sources/sets of specificaitons</u>** (e.g., claims and EHRs), section **2b6** also must be completed.
- Respond to <u>all</u> 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.

<u>Note</u>: 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: Click here to describe

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*).

Medicare 100% Standard Analytic File (SAF), Inpatient file and Denominator file.

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

4Q CY2012 – 1Q CY2016

1.4. What levels of analysis were tested? (*testing must be provided for <u>all</u> 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 <u>measured entities</u> 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*)

Testing of the measure, as currently specified, was performed at the facility level using the Medicare 100% SAF covering 4Q CY2012 through 1Q CY2016 claims. The Inpatient file containing information on inpatient claims, and the Denominator file containing information on the enrollment and demographics were used for purposes of testing and analysis. We included patients with an eligible index admission (denominator) between 1Q CY2013 and 4Q CY2015. An additional quarter of data (4Q CY2012) was included in the analysis to ensure that all risk adjustment factors were accurate. An additional quarter of data (1Q CY2016) was also included in the analysis to ensure that all eligible unplanned readmissions were captured. This dataset was used without modification for the following testing activities: generating performance rates and descriptive statistics for the measure submission; evaluating measure exclusions; and, evaluating sociodemographic (SDS) variables for potential inclusion.

A total of 4,974 short-term acute care hospitals were included. Short-term acute care hospitals were defined as: PPS-Exempt Cancer Hospitals (PCH); short-term acute care Prospective Payment System (PPS) hospitals; and, Critical Access Hospitals (CAH). Hospitals from Maryland were included in this analysis. All other acute hospital types were excluded, such as: Long-Term Care Hospitals (LTCHs), Inpatient Rehabilitation Facilities

(IRFs), and Inpatient Psychiatric Hospitals. The hospital was the level of analysis, as defined by National Provider Identifier (NPI).

Below are descriptive statistics for the 4,974 short-term acute care hospitals included in the measure denominator:

	All Short-Term Acute Care Hospitals	PPS-Exempt Cancer Hospitals	Short-Term Acute Care PPS Hospitals	Critical Access Hospitals
Number of Hospitals	4,974	11	3,617	1,346
% of Hospitals in Full Dataset	*100.00%*	*0.22%*	*72.72%*	*27.06%*
Number of Admissions (Denominator)	3,067,675	73,159	2,934,917	59,599
% of Admissions in Full Dataset	*100.00%*	*2.38%*	*95.67%*	*1.94%*
Mean Admissions per Hospital (Standard Deviation)	616.74 (1,151.06)	6,650.82 (7,861.36)	811.42 (1,174.69)	44.28 (42.89)
Range (Min-Max)	1-22,300	291-22,300	1-12,998	1-308
Quartile Range	713	7,242	951	951
Minimum	1	291	1	1
25th percentile	37	1,071	107	13
50th percentile	159	3,364	388	31
75th percentile	750	8,313	1,058	61
Maximum	22,300	22,300	12,998	308

Table 1: Shows unadjusted denominator population (1Q2013-4Q2015) for short-term acute care hospitals (i.e., PCHs, short-term acute care PPS hospitals, and CAHs) included in measure testing for the *30-Day Unplanned Readmissions for Cancer Patients* measure. Data source: Analysis of Medicare SAF (4Q2012-1Q2016), based on data provided by Watson Policy Analysis, 01/13/2017.

Exceptions

- Reliability testing of the performance measure score used a subset of the dataset above, as described in Section 1.7.
- Empirical validity testing of the performance measure score used a subset of the dataset above, as described in Section 1.7.
- Risk adjustment testing of the performance measure score used a development and a split sample of the dataset above, as described in Section 2b4.5.
- Evaluating meaningful differences of the performance scores used a subset of the dataset above, as described in Section 2b5.1.
- In 2015, we examined the validity of the Type of Admission/Visit reported via the *UB-04 Uniform Bill Locator 14* (Claim Inpatient Admission Type Code in the Medicare SAF) to accurately identify planned and unplanned readmissions in this measure. We summarize the results of this data element validity testing in Sections 2b2.2-2b2.4 of testing attachment. This testing was performed using a mix of administrative claims data and manually-abstracted data. For simplicity purposes, this dataset is not described in this testing attachment. A complete description of the testing dataset and testing results is included in the Appendix.

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*)

Testing and analysis were performed at the facility level. All testing activities were based on 100% of claims with no sampling applied, excepted as noted above. Below are descriptive statistics for the patients included in the measure denominator and numerator, along with unadjusted *30-Day Unplanned Readmissions for Cancer Patients* rates:

Variable	Value	Denominator		Numerator		*30-Day Unplanned Readmission Rate-Unadjusted*
		N	% of Total	N	% of Total	
Total		3,067,675	100.00%	587,915	100.00%	19.16%
Sex	Unknown	5,560	0.18%	1,043	0.18%	18.76%
	Male	1,616,259	52.69%	317,592	54.02%	19.65%
	Female	1,445,856	47.13%	269,280	45.80%	18.62%
Race	Unknown	28,494	0.93%	5,328	0.91%	18.70%
	White	2,535,852	82.66%	469,232	79.81%	18.50%
	Black	354,140	11.54%	82,120	13.97%	23.19%
	Other	39,428	1.29%	8,062	1.37%	20.45%
	Asian	42,990	1.40%	9,087	1.55%	21.14%
	Hispanic	52,158	1.70%	11,182	1.90%	21.44%
	North American Native	14,613	0.48%	2,904	0.49%	19.87%
Age	Unknown	5,560	0.18%	1,043	0.18%	18.76%
	Under 65	409,844	13.36%	95,759	16.29%	23.36%
	65-69	618,508	20.16%	116,829	19.87%	18.89%
	70-74	606,147	19.76%	116,729	19.85%	19.26%
	75-79	529,837	17.27%	102,143	17.37%	19.28%
	80-84	424,681	13.84%	78,297	13.32%	18.44%
	85+	473,098	15.42%	77,115	13.12%	16.30%
Dual-Eligible Status	Never Dual Eligible	2,448,890	79.83%	448,721	76.32%	18.32%
	Dual Eligible at Some Point	618,785	20.17%	139,194	23.68%	22.49%

<u>Table 2</u>: Includes the distribution of four patient-level variables for the measure denominator and numerator populations (along with unadjusted *30-Day Unplanned Readmissions for Cancer Patients* rates) when the *30-Day Unplanned Readmissions for Cancer Patients* measure is applied to 1Q CY2013-4Q CY2015 index admissions for all 4,974 short-term acute care hospitals (defined as PCHs, short-

term acute care PPS hospitals, and CAHs). Data source: Analysis of Medicare SAF (4Q2012-1Q2016), based on data provided by Watson Policy Analysis, 01/13/2017. Variables are defined as follows:

- "Sex" variable—defined in the Medicare SAF;
- "Race" variable—defined in the Medicare SAF;
- "Age" variable—populated by adding one year to the "Age" field in the Medicare SAF (1Q2013-1Q2016), which is reported as the beneficiary's age at the end of the prior year; and,
- "Dual-Eligible Status" variable—used as a proxy for socioeconomic status and is populated by analyzing the Buyin field in the Medicare SAF (1Q2013-1Q2016). Patients with any claims with a value of "A", "B", or "C" in the Buyin field in the 1Q2013-1Q2016 dataset are coded as "Dual-Eligible" in this variable. All other patients are coded as "Never Dual-Eligible" in this variable.

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.

Performance Score Reliability Dataset

Reliability of the performance score was tested using a subset of the dataset described in Section 1.5. First, we limited the dataset to hospitals with a minimum of 50 eligible index admissions for the CY2013-CY2015 period. Then, we randomly split the dataset into two equal and distinct patient subsets to calculate risk-adjusted performance scores at the hospital level. Finally, we compared the performance scores for each hospital. We then repeated the analysis by year for CY2013, CY2014, and CY2015 to examine measure stability over time.

Below is summary-level information for these datasets, with the minimum case count of 50 eligible index admissions applied:

	2013-15	2013	2014	2015
Sample Size Cutoff: 50 Index Admissions/Hospital (25/Sample Set)				
Number of Hospitals	3,502	2,575	2,559	2,511
% of Hospitals in Full Testing Dataset	*70.41%*	*54.37%*	*54.19%*	*53.56%*
Number of Admissions (Denominator)	3,038,015	1,000,165	980,725	977,975
% of Admissions in Full Testing Dataset	*99.03%*	*96.36%*	*96.50%*	*96.50%*

Table 3: Includes summary-level information for the datasets used for performance score reliability testing for the *30-Day Unplanned Readmissions for Cancer Patients* measure when applied to 1Q CY2013-4Q CY2015 index admissions for short-term acute care hospitals (defined as PCHs, short-term acute care PPS hospitals, and CAHs). The dataset was limited to hospitals with a minimum of 50 eligible index admissions (in total and by year), then randomly split into two equal subsets. Data source: Analysis of Medicare SAF (4Q2012-1Q2016), based on data provided by Watson Policy Analysis, 01/13/2017.

Performance Score Empirical Validity Dataset

Empirical validity of the performance score was tested using a subset of the dataset described in Section 1.5. We limited the dataset to eligible index admissions for the 3Q CY2014-2Q CY2015 period. Below is summary-level information for this dataset:

	3Q CY2014-2Q CY2015
Number of Hospitals	4,720
% of Hospitals in Full Testing Dataset	*94.89%*
Number of Admissions (Denominator)	1,018,500

% of Admissions in Full Testing Dataset	*33.20%*

Table 4: Includes summary-level information for the dataset used for performance score empirical validity testing for the *30-Day Unplanned Readmissions for Cancer Patients* measure when applied to 1Q CY2013-4Q CY2015 index admissions for short-term acute care hospitals (defined as PCHs, short-term acute care PPS hospitals, and CAHs). The dataset was limited to hospitals with eligible index admissions in 3Q2014-2Q2015. Data source: Analysis of Medicare SAF (4Q2012-1Q2016), based on data provided by Watson Policy Analysis, 01/13/2017.

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).

We considered the NQF's guidelines for patient-level SDS risk adjustment in evaluating potential risk factors for our measure. We found that only "Race" and "Dual-Eligible Status" were readily available in the dataset we used for measure testing. We selected both variables for potential inclusion in our risk adjustment model. Several peer-review publications have suggested disparities in readmissions across conditions [1-5] and in cancer,[6-13] though the results were at times conflicting or not significant.

We utilized the "Race" variable, as defined in the Medicare SAF. The "Dual-Eligible Status" variable was used as a proxy for socioeconomic status and was populated by analyzing the Buyin field in the Medicare SAF. Patients with any claims with a value of "A" ("Part A, State Buy-In"), "B" ("Part A, State Buy-In"), or "C" ("Parts A and B, State Buy-in") in the Buyin field are coded as "Dual-Eligible" in this variable. All other patients are coded as "Never Dual-Eligible" in this variable.

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*)

In developing this measure, we aimed to produce a reliable measure that yielded similar repeat measurements for each facility and that is stable over time. Accordingly, we evaluated the measure's ability to generate consistent results through split-half correlations when randomly selected subsets of patients were measured in total and over time. Through this test-retest approach, we randomly split the dataset into two equal and distinct patient subsets to calculate risk-adjusted performance scores at the hospital level. As metrics of agreement, we calculated both the Intraclass Correlation Coefficient (ICC) and Spearman-Brown Prophecy Formula (S-B) for each of the iterations.[14] The ICC is estimated from the random effects model, which produces risk-adjusted rates.[15] The S-B formula effectively projects the correlation as if the full sample was used and not split randomly. This analysis was conducted over 100 iterations to evaluate the measure's reliability using CY2013-CY2015 eligible index admissions, Further, the S-B value was calculated for each year for the unadjusted data which allowed us to assess any bias that may occur from randomly splitting the data.

By combining multiple years of data, we were able to include more cases to confirm the measure's reliability. We tested the hypothesis that the S-B statistics from each year were greater than 0.5, indicating strong reliability and large effect size.[16,17] Confidence intervals based on the 100 simulations were calculated for the unadjusted and risk-adjusted rates based on the split-half samples.

Because hospitals with fewer cases were expected to have less reliable estimates, we established a minimum volume threshold to reduce potential "noise" associated with calculating performance rates for smaller-volume hospitals. We modeled minimum case counts of 22, 50, and 75 index admissions per hospital (in total and by year) in performing the split half correlation analysis. We found that a minimum case count of 50 index admissions (25 per subset) per hospital produced consistent and stable results, while limiting the number of hospitals excluded from the analysis. Therefore, we limited reliability testing of performance scores to hospitals with a minimum of 50 eligible index admissions during each measurement period—in total and then by year for CY2013, CY2014, and CY2015.

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*)

In applying the test-retest approach to establish measure reliability, we set a minimum case count of 50 index admissions (25 per subset) per hospital. There were 3,502 facilities included in the 100 split-half simulations for CY2013-CY2015. For the three-year period, the ICC for the unadjusted rates was 0.570 (95% CI: 0.567, 0.572), while the ICC for the risk-adjusted rates was 0.482 (95% CI: 0.479, 0.485).

The S-B statistic allows us to project what the reliability would be if the entire sample were used instead of the split sample. For the three-year period (CY2013-CY2015), the mean S-B was 0.726 (95% CI: 0.724, 0.728) and 0.650 (95% CI: 0.648, 0.653) for unadjusted and risk-adjusted values, respectively.

We also examined the stability of the measure, testing the hypothesis that our mean S-B from each year are greater than 0.5. We applied a minimum case count of 50 index admissions (25 per subset) per hospital for each year. Below are the results of that testing, including S-B statistics for unadjusted and risk-adjusted rates:

	2013	2014	2015
Sample Size Cutoff: 50 Index Admissions/Hospital (25/Sample Set)			
Number of Hospitals	2,575	2,559	2,511
Number of Admissions (Denominator)	1,000,165	980,725	977,975
Unadjusted Rates			
Mean S-B Score	0.635 (0.012)	0.620 (0.014)	0.608 (0.013)
95% Confidence Interval	(0.632, 0.637)	(0.618, 0.623)	(0.605, 0.610)
t-test of H_0: S-B <= 0.5 (p-value)	113.98 (<0.001)	87.32 (<0.001)	80.18 (<0.001)
Risk-Adjusted Rates			
Mean S-B Score (Standard Deviation)	0.543 (0.015)	0.530 (0.017)	0.502 (0.017)
95% Confidence Interval	(0.540, 0.546)	(0.526, 0.533)	(0.499, 0.506)
t-test of H_0: S-B <= 0.5 (p-value)	29.31 (<0.001)	17.62 (<0.001)	1.281 (0.203)

Table 5: Shows statistical results of reliability testing by year for the *30-Day Unplanned Readmissions for Cancer Patients* measure when applied to 1Q CY2013-4Q CY2015 index admissions for short-term acute care hospitals (defined as PCHs, short-term acute care

PPS hospitals, and CAHs). Includes mean S-B scores with standard deviation and 95% confidence interval along with measures of significance (t-test and p-value), when applied to unadjusted and risk-adjusted rates. Each dataset was limited to hospitals with a minimum of 50 eligible index admissions (in total and by year). Data source: Analysis of Medicare SAF (4Q2012-1Q2016), based on data provided by Watson Policy Analysis, 01/13/2017.

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?*)

When examining the reliability testing results for the three-year period (CY2013-CY2015) and the minimum case count of 50 index admissions (25 per subset) per hospital, we observed reliability scores that could be interpreted as "fair" or "strong" depending on the statistical test. The ICC was 0.570 (95% CI: 0.567, 0.572) and 0.482 (95% CI: 0.479, 0.485), for unadjusted and risk-adjusted values, respectively. This result may be interpreted as "fair" reliability.[18] The mean S-B for the same period was 0.726 (95% CI: 0.724, 0.728) for unadjusted rates and 0.650 (95% CI: 0.648, 0.653) for risk-adjusted rates. Both of these values are significantly higher than the 0.5 that indicates a large effect size with p-values < 0.001.[16,17]

Similarly, the S-B values for each year individually (CY2013, CY2014, and CY2015) demonstrated that the reliability measures were stable and consistent over time, producing bell-shaped distributions across all years. The S-B analysis of unadjusted rates by year generated mean S-B exceeding 0.60 with p-values < 0.001, indicating strong reliability and large effect size based on accepted conventional interpretation. When applied to risk-adjusted rates, the S-B analysis generated mean S-B values exceeding 0.50 (p-values < 0.001) in 2013 and 2014, indicating strong reliability and large effect size based on accepted conventional interpretation.[16,17] Overall, the consistent calculations between the two data randomly-split subsets for each period provided evidence that performance variations between hospitals were attributable to hospital-level factors, rather than patient-level factors.

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)*

Data Element Validity

In developing this measure, we aimed to utilized data elements within administrative claims data that have face validity and reliability. Previous measure testing utilized manual chart review to assess the validity of the Type of Admission/Visit reported via the *UB-04 Uniform Bill Locator 14* (Claim Inpatient Admission Type Code in the Medicare SAF) to accurately identify planned and unplanned readmissions. Sensitivity and specificity of the claims-based indicator were evaluated across the participating facilities and in the aggregate to establish data element validity. Please see the Appendix for testing details and results.

Performance Measure Score Empirical Validity Testing

As a test of empirical validity, similar facility-level performance should be observed among measures that evaluate similar healthcare processes. Thus, in testing this measure, we aimed to compare the relative

performance of hospitals under the *30-Day Unplanned Readmissions for Cancer Patients* measure with another measure that is conceptually related. Due to significant cancer measurement gaps, we did not identify NQF-endorsed cancer-specific process or outcome measures suitable for this purpose. However, we determined that CMS' *Hospital-Wide All-Cause Readmission Measure (HWR)* (NQF #1789) could be used for empirical validity testing.

While the two measures have different target populations, they both utilize Medicare claims administrative claims data and assess unplanned readmissions within thirty days of hospital discharge. Additionally, the *30-Day Unplanned Readmissions for Cancer Patients* measure was modeled after the *HWR (NQF #1789)* measure where possible (e.g., a readmission can be counted as both as readmission and an eligible index admission under both measures). Third, both measures are adjusted for patient-level risk factors. Finally, performance score reliability of both measures has been established through measure testing. Thus, within each measure, performance variations between hospitals can be attributed to hospital-level factors (e.g., practice patterns that lead to treatment complications, inadequate discharge planning, comorbidity management, and follow-up care) rather than patient-level factors.

The hypothesized relationship is that better performance (i.e., lower hospital-level rates) on the *HWR (NQF #1789)* measure should be associated with better performance (i.e., lower hospital-level rates) on the *30-Day Unplanned Readmissions for Cancer Patients* measure. To test this hypothesis, we compared hospital-level performance rates for both measures for the 3Q2014-2Q2015 period, the latest period for which *HWR (NQF #1789)* rates are available on Hospital Compare. We calculated the correlation coefficient between rates as an indicator of the strength of the associations. Moderate positive correlation is expected, given that the measures assess similar healthcare practices related to patient care. We limited this analysis to short-term acute care PPS hospitals, as *HWR (NQF #1789)* rates are not reported on Hospital Compare for PCHs and CAHs.

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

Data Element Validity
Previous data element validity testing of the Type of Admission/Visit reported via the UB-04 Uniform Bill Locator 14 (Claim Inpatient Admission Type Code in the Medicare SAF), which utilized manual chart review, generated global sensitivity and specificity scores of 0.879 and 0.896, respectively.

Performance Measure Score Empirical Validity Testing
An overall correlation of 0.2769 ($p<0.001$) was observed for the 4,719 hospitals included in the analysis.

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?*)

Data Element Validity
Global sensitivity and specificity scores of 0.879 and 0.896, respectively, confirmed the validity of the Type of Admission/Visit reported via the *UB-04 Uniform Bill Locator 14* (Claim Inpatient Admission Type Code in the Medicare SAF) to accurately identify planned and unplanned readmissions, as validated by chart review.

Performance Measure Score Empirical Validity Testing
As expected, we observed a statistically significant, moderate positive correlation (0.2769) between the *30-Day Unplanned Readmissions for Cancer Patients* measure and the *HWR (NQF #1789)* measure. This confirms our hypothesis that better performance on the *HWR (NQF #1789)* measure is associated with better performance on the *30-Day Unplanned Readmissions for Cancer Patients* measure.

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*)

In developing this measure, we aimed to establish a cohort of patients with cancer that have unplanned readmissions due to potentially avoidable treatment complications or inadequate discharge instructions or coordination with follow-up outpatient care. Thus, for the denominator population, we applied the following exclusions for patients: 1) Less than 18 years of age; 2) who died during the index admission; 3) discharged Against Medical Device (AMA); 4) transferred to another acute care hospital; 5) discharged with a planned readmission; 6) having missing or incomplete data; and, 7) not admitted to an inpatient bed.
Readmissions for patients with progression of disease (using a principal diagnosis of metastatic disease as a proxy) and for patients with planned admissions for treatment (defined as a principal diagnosis of chemotherapy or radiation therapy) are excluded from the measure numerator.

All exclusions were determined by careful review with our workgroup along with guidance from data and coding experts. They reflect clinically-relevant decisions and alignment with coding practices to ensure accurate performance rates. We examined the frequency of the exclusions to assess their impact on the measurement cohort and performance rates using the dataset described in Sections 1.2-1.6. The rationale for each numerator exclusion is provided in Section S.5 (Numerator Details) of the Measure Information Form. The rationale for each denominator exclusion is provided in Section S.9 (Denominator Exclusion Details) of the Measure Information Form.

To assess the potential impact of each exclusion, we examined the overall number and percentage of each denominator and numerator exclusion in our dataset.

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*)

To assess the potential impact of each exclusion, we examined the overall frequency of each denominator and numerator exclusion in our dataset. The diagram below shows the overall number and percentage of eligible index admissions excluded from the denominator:

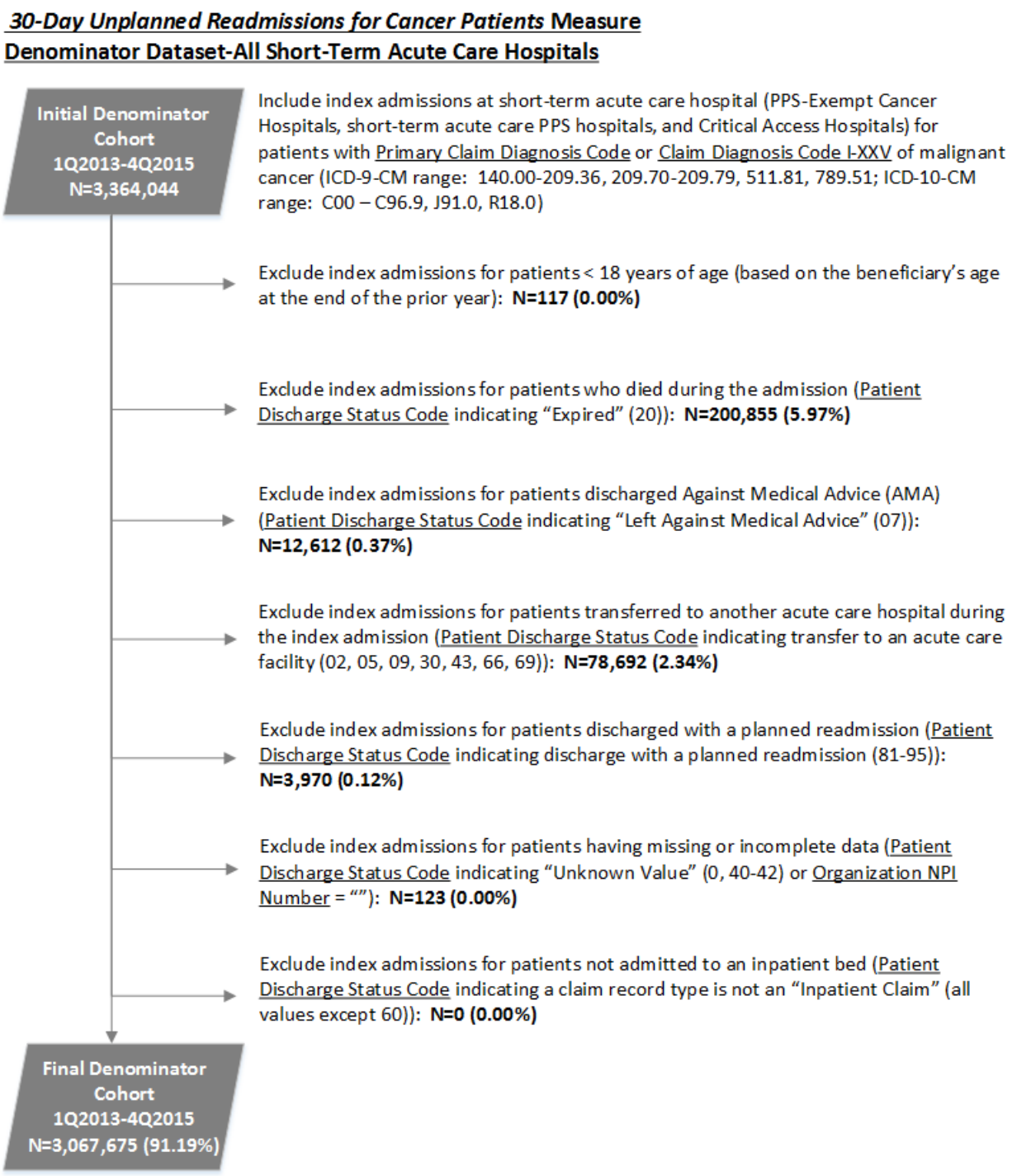


Figure 1: Shows the overall number and distribution of each denominator exclusion for the *30-Day Unplanned Readmissions for Cancer Patients* measure when applied to 1Q CY2013-4Q CY2015 index admissions for short-term acute care hospitals (defined as PCHs, short-term acute care PPS hospitals, and CAHs). Data source: Analysis of Medicare SAF (4Q2012-1Q2016), based on data provided by Watson Policy Analysis, 01/13/2017.

The diagram below shows the overall number and percentage of eligible index admissions excluded from the denominator:

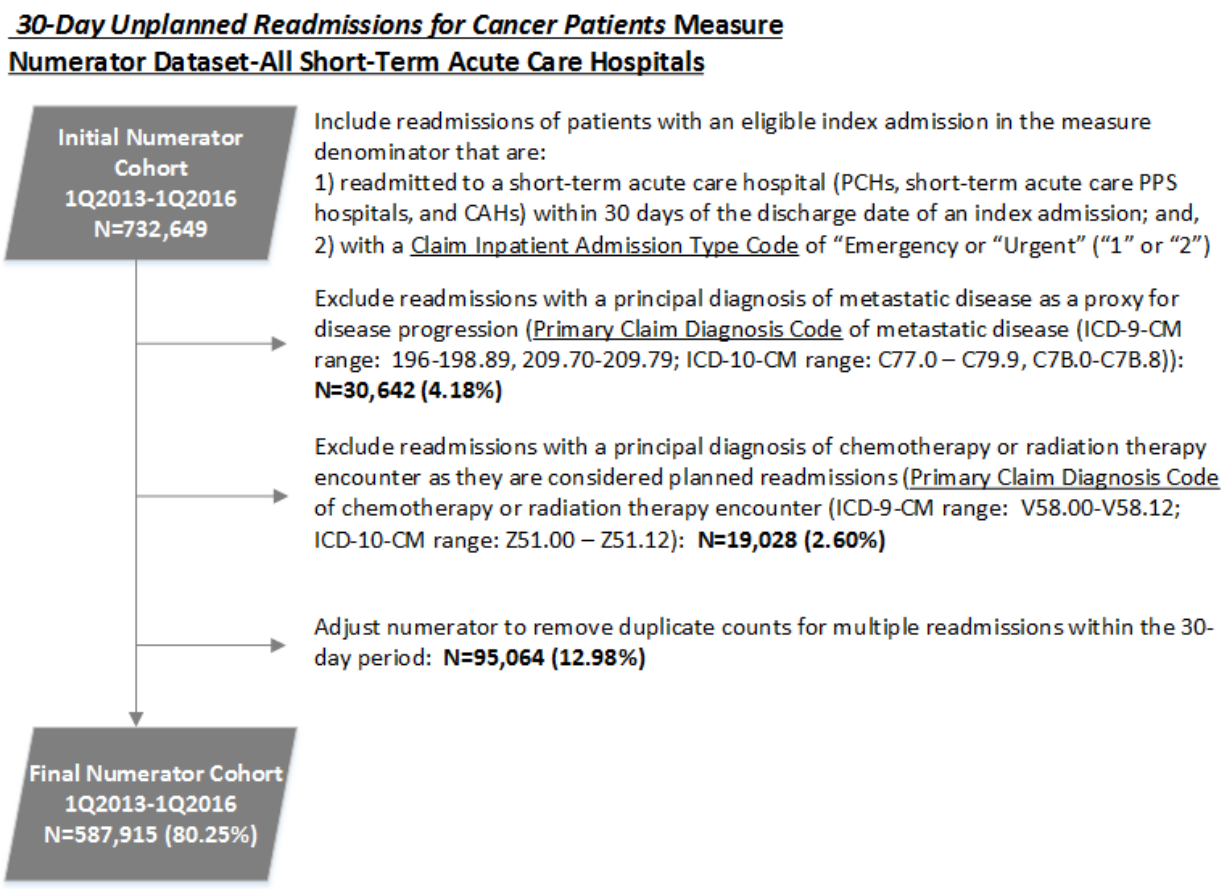


Figure 2: Shows the overall number and distribution of each numerator exclusion for the *30-Day Unplanned Readmissions for Cancer Patients* measure when applied to 1Q CY2013-4Q CY2015 index admissions for short-term acute care hospitals (defined as PCHs, short-term acute care PPS hospitals, and CAHs). Data source: Analysis of Medicare SAF (4Q2012-1Q2016), based on data provided by Watson Policy Analysis, 01/13/2017.

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*)

Denominator Exclusions

The overall frequency of the denominator exclusions is low, reducing the initial denominator cohort by less than 10%. Patients who died during the index admission yielded the highest exclusions (N=200,855, 5.97%), while patients not admitted to an inpatient bed yielded no exclusions. Based on our review of these exclusions, we think they should be retained for the following reasons:

- Age less than 18 years of age, based on the beneficiary's age at the end of the prior year.

  Rationale: Pediatric patients represent a very small and distinct Medicare population with different characteristics and outcomes.

- Patients who died during the index admission, defined as *Patient Discharge Status Code* indicating "Expired" (20).

  Rationale: Patients that die during the index admission cannot be readmitted.

- Patients discharged AMA, defined as *Patient Discharge Status Code* indicating "Left Against Medical Advice" (07).

Rationale: The hospital had limited opportunity to ensure the patient was prepared for discharge and had appropriate follow-up care.

- Patients transferred to another acute care hospital during the index admission, defined as *Patient Discharge Status Code* indicating transfer to an acute care facility (02, 05, 09, 30, 43, 66, 69).

  Rationale: Responsibility for any unplanned readmissions is assigned to the final discharging hospital. Intermediate index admissions within a single episode of care are ineligible for inclusion.

- Patients discharged with a planned readmission, defined as *Patient Discharge Status Code* indicating discharge with a planned readmission (81-95).

  Rationale: The patient was discharged with a planned readmission, which is ineligible for the measure numerator.

- Patients having missing or incomplete data, defined as *Patient Discharge Status Code* indicating "Unknown Value" (0, 40-42) or Organization NPI Number = "".

  Rationale: Admissions without a valid discharge status cannot be evaluated for measure exclusions. Admissions with a discharge status reserved for hospice claims only are not admissions for acute care or to acute care hospitals. Claims without an Organizational NPI Number cannot be evaluated for inclusion in the measure.

- Patients not admitted to an inpatient bed, defined as *NCH Claim Type Code* indicating a claim record type is not an "Inpatient Claim" (all values except 60).

  Rationale: These admissions are not for acute care or to acute care hospitals.

Numerator Exclusions

The overall frequency of the numerator exclusions is low, reducing the initial numerator cohort by less than 7%. Patients with readmissions for progression of disease yielded the highest exclusions (N=30,642, 4.18%), while patients with planned admissions for treatment yielded the lowest exclusions (N=19,028, 2.60%). Based on our review of these exclusions, we think they should be retained for the following reasons:

- Readmissions for patients with progression of disease, defined as *Primary Claim Diagnosis Code* of metastatic disease (ICD-9-CM range: 196-198.89, 209.70-209.79; ICD-10-CM range: C77.0 – C79.9, C7B.0-C7B.8).

  Rationale: A primary (or principal) diagnosis of metastatic disease serves as a proxy for disease progression. Readmissions for conditions or symptoms associated with disease progression are not reflective of poor clinical care but, rather, advanced disease.

- Readmissions for patients with planned admissions for treatment, defined as *Primary Claim Diagnosis Code* of chemotherapy or radiation encounter (ICD-9-CM range: V58.00-V58.12; ICD-10-CM range: Z51.00 – Z51.12).

  Rationale: Readmissions are expected and planned for some patients who require additional cancer treatment in the inpatient setting. These readmissions reflects high-quality care that is focused on patient safety and are reliably distinguishable in claims data.

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 11 risk factors with 15 values**

☐ **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.

Statistical Risk Model Method

We developed a statistical risk model based on a comparison of observed vs. expected rates, as is commonly used in healthcare outcomes. We used logistic regression to estimate the probability of an unplanned readmission, based on the measure specifications and risk factors described herein. The probability of unplanned readmission was then summed over the index admissions for each hospital to calculate the *expected* unplanned readmission rate. We then summed the actual or *observed* unplanned readmissions for each hospital and calculated the ratio of *observed* unplanned readmissions to *expected* unplanned readmissions for each hospital. Each hospital's ratio was then multiplied by the national or standard unplanned readmissions rate to generate the risk-adjusted *30-Day Unplanned Readmissions for Cancer Patients* rate (see formula below). Lower risk-adjusted rates (observed/expected ratios) are interpreted as better quality while higher risk-adjusted rates (observed/expected ratios) indicate poorer quality.

$$Risk - Adjusted\ Rate = \frac{observed\ rate}{expected\ rate} \times national\ or\ standard\ rate$$

Figure 3: Risk-adjusted rate formula for the *30-Day Unplanned Readmissions for Cancer Patients* measure. A lower observed/expected ratio is interpreted as better quality, while a higher ratio indicates poorer quality.

Risk Factors with Coefficients

Below are the risk factors included in the risk adjustment methodology, with coefficients and odds ratio estimates:

Parameter-redo all numbers	Model Coefficients		Odds Ratio Estimates		
	Estimate	P-Value	Point Estimate	95% Wald Confidence Limits	
Intercept	-2.966	<.0001			
ICU Stay	0.055	<.0001	1.117	1.106	1.127
Male	0.046	<.0001	1.097	1.088	1.106
Dual-Eligible Status	0.069	<.0001	1.147	1.135	1.159
Surgical Admission	-0.226	<.0001	0.637	0.631	0.643
Multiple Comorbidities	0.123	<.0001	1.279	1.266	1.293
Solid Tumor *(excluding Metastatic Disease)*	-0.079	<.0001	0.854	0.847	0.861
Length of Stay Greater than 3 Days	0.149	<.0001	1.347	1.335	1.360
Age: < 65	*Reference Age*				
Age: 65-69	-0.075	<.0001	0.861	0.849	0.874
Age: 70-74	-0.068	<.0001	0.873	0.860	0.885
Age: 75-79	-0.078	<.0001	0.856	0.844	0.869
Age: 80-84	-0.101	<.0001	0.818	0.805	0.831
Age: 85+	-0.162	<.0001	0.723	0.712	0.735
Hospitalization in the Prior 60 Days	0.239	<.0001	1.612	1.597	1.627
Discharged to Home	-0.109	<.0001	0.804	0.797	0.811
Discharged to Hospice	-1.277	<.0001	0.078	0.075	0.080

Table 6: Shows model coefficients and odds ratio estimates for risk variables included in the risk adjustment model for the *30-Day Unplanned Readmissions for Cancer Patients* measure. Data source: Analysis of Medicare SAF (4Q2012-1Q2016), based on data provided by Watson Policy Analysis, 01/13/2017.

Risk Factor Definitions and Codes

Below are the definitions and code lists (where applicable) for the risk factors included in the logistic regression model:

- ICU Stay: Index admissions for patients with an ICU stay during the index admission, as indicated by a *Revenue Center Code* (Code Range: 0200-0209) in the Medicare SAF, are coded as "1" in this variable. All other index admissions are coded as "0" in this variable. *Please see the data dictionary for the complete code list with descriptions.*
- Male: Index admissions for patients listed as "Male" in the "Sex" field in the Medicare SAF, are coded as "1" in this variable. All other index admissions are coded as "0" in this variable.
- Dual-Eligible Status: Index admissions for dual-eligible Medicare beneficiaries, as indicated by a value of "A", "B", or "C" in the Buyin field in the Medicare SAF, are coded as "1" in this variable. All other index admissions are coded as "0" in this variable. *Please see the data dictionary for the complete code list with descriptions.*
- Surgical Admission: Index admissions for patients that had surgery during the index admission, as indicated by a surgical Medicare Severity-Diagnosis Related Groups (MS-DRG) in the *Claim Diagnosis Related Group Code* in the Medicare SAF, are coded as "1" in this variable. All other index admissions are coded as "0" in

this variable. Surgical MS-DRGs are commonly used in the Agency for Healthcare Research and Quality (AHRQ) Patient Safety Indicators (PSI) and, therefore, were used in this risk adjustment model.[19]

- Multiple Comorbidities: Index admissions for patients with 2 or more comorbidities, as defined within the Elixhauser Comorbidity Index and indicated by a corresponding *Primary Claim Diagnosis Code* and *Claim Diagnosis Code I-XXV*, are coded as "1" in this variable. The Elixhauser comorbidity categories are commonly used in AHRQ PSIs and, therefore, were used in this risk adjustment model.[19] We excluded from the comorbidity count the comorbidities for "Tumor," "Lymph," and "Mets" since they would be common in the cancer population and a separate indicator variable was constructed for metastatic disease. All other index admissions are coded as "0" in this variable. *Please see the https://www.hcup-us.ahrq.gov/toolssoftware/comorbidity/comformat2012-2015.txt for the complete code list with descriptions.*
- Solid Tumor *(Excluding Metastatic Disease)*: Index admissions for patients with non-metastatic, non-hematologic cancer, as indicated by a *Primary Claim Diagnosis Code* or *Claim Diagnosis Code I-XXV* (ICD-9-CM range: 140.00-195.89, 199.00-199.99, 209.00-209.36, 511.81, 789.51; ICD-10-CM range: C00.0 – C76.9, C80.0-C80.2, J91.0, R18.0) in the Medicare SAF, are coded as "1" in this variable. All other index admissions are coded as "0" in this variable. *Please see the data dictionary for the complete code list with descriptions.*
- Length of Stay Greater than 3 Days*:* Index admissions for patients with a length of stay greater than 3 days during the index admission, calculated as *Claim Through Date* – *Claim Admission Date* ≥ 3 in the Medicare SAF, are coded as "1" in this variable. All other index admissions are coded as "0" in this variable.
- Age < 65: Index admissions for patients aged < 65 years at the end of the prior year, based on the "Age" field in the Medicare SAF, are coded as "1" in this variable. All other index admissions are coded as "0" in this variable. **Note, Age < 65 is included in the risk adjustment model as the baseline age category or reference point only.*
- Age 65-69: Index admissions for patients aged 65-69 years at the end of the prior year, based on the "Age" field in the Medicare SAF, are coded as "1" in this variable. All other index admissions are coded as "0" in this variable.
- Age 70-74: Index admissions for patients aged 70-74 years at the end of the prior year, based on the "Age" field in the Medicare SAF, are coded as "1" in this variable. All other index admissions are coded as "0" in this variable.
- Age 75-79: Index admissions for patients aged 75-79 years at the end of the prior year, based on the "Age" field in the Medicare SAF, are coded as "1" in this variable. All other index admissions are coded as "0" in this variable.
- Age 80-84: Index admissions for patients aged 80-84 years at the end of the prior year, based on the "Age" field in the Medicare SAF, are coded as "1" in this variable. All other index admissions are coded as "0" in this variable.
- Age 85+: Index admissions for patients aged 85 years and older at the end of the prior year, based on the "Age" field in the Medicare SAF, are coded as "1" in this variable. All other index admissions are coded as "0" in this variable.
- Hospitalization in the Prior 60 Days: Index admissions for patients discharged from a hospitalization at a short-term acute care hospital within 60 days of the admission date of the index admission, calculated as *Claim Through Date* (of any prior hospitalizations) – *Claim Admission Date* (of the index admission) ≤ 60 in the Medicare SAF, are coded as "1" in this variable. All other index admissions are coded as "0" in this variable.

- Discharged to Home: Index admissions for patients that are discharged to home/self care, as indicated by a value of "01" in the *Patient Discharge Status Code* in the Medicare SAF, are coded as "1" in this variable. All other index admissions are coded as "0" in this variable.
- Discharged to Hospice: Index admissions for patients that are discharged to hospice – home or hospice – medical facility, as indicated by a value of "50" or "51"in the *Patient Discharge Status Code* in the Medicare SAF, are coded as "1" in this variable. All other index admissions are coded as "0" in this variable.

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*)

Conceptual/Clinical Review
We identified potential risk factors for the *30-Day Unplanned Readmissions for Cancer Patients* measure using the following methods:
- Review of the literature to determine which patient-level risk adjustors were included in risk- adjusted NQF-endorsed measures and measures included in CMS public reporting programs; and,
- Convening a multidisciplinary workgroup of:
  - Physician subject-matter experts from cancer hospitals to identify patient-level risk adjustors that are clinically-relevant for unplanned readmissions in patients with cancer;
  - Data analysts with experience in complex analyses of hospital data, quality measurement, and quality improvement, with a specific focus on cancer conditions;
  - Experienced coders to advise on the selection and completeness of code lists for the measure; and,
  - Analytics experts with experience in statistical testing methods and in creating predictive models for unplanned readmissions.

In total, 25 patient-level variables were evaluated for potential inclusion in the risk adjustment model. The list of potential risk adjustors was then compared to the data elements available in administrative claims data. Since this measure is to be implemented using claims data only, 7 clinical and SDS variables (Table 7, Group A) that are not well-defined in claims data were not included in this model. Additionally, 2 variables (Table 7, Group B) were unavailable in our measure testing dataset. The list of potential risk adjustors was then refined to include only variables *not* in the control of the hospital, as the goal of this model is to adjust for patient-specific factors only. This eliminated 1 variable (Table 7, Group C). Finally, 1 SDS variable ("Race") was removed (Table 7, Group D). While there is evidence that racial minorities have higher readmission rates, the evidence is, at times, conflicting or non-significant.[1,3,4,6-12] Moreover, Joynt, et al. found that racial disparities in readmissions were related to patient race *and* the site of care, suggesting an opportunity to reduce disparities in care.[2] In view of these findings and because we could not articulate a causal relationship between race and readmission, we removed the variable to ensure that the risk adjustment model would not mask disparities in care. This process yielded 14 risk factors (Table 7, Group E) to be evaluated for fit in the risk adjustment model. Throughout this process, all potential risk factors were determined by careful review with workgroup members. They reflect clinically-relevant decisions and alignment with coding practices and analytical standards to ensure accurate assessments of patient-level risk factors present at the index admission and outside the control of the hospital.

Below is the complete list of potential risk factors identified through the workgroup's review, with the workgroup's assessment:

Potential Risk Adjustors Evaluated for This Measure
A. Risk Factors Removed from Consideration-*Not Well-Defined in Claims Data* (7)
Severity of Illness
Local vs. Regional vs. Distant Disease
High-Risk Medication Use
Psychological Services
Early Palliative Care/Hospice
History of Substance Abuse
Nutritional Status
B. Risk Factors Removed from Consideration-*Not Available in Measure Testing Dataset* (2)
Marital Status
Geographic Distance from Hospital
C. Risk Factors Removed from Consideration-*Within Hospital's Control* (1)
Weekday vs. Weekend Discharge
D. Risk Factors Removed from Consideration-*Potential to Mask Disparities in Care* (1)
Race
E. Risk Factors Evaluated for Fit in the Risk Adjustment Model (14)
Age
Gender
Dual-Eligible Status *(Proxy for Socioeconomic Status)*
Number of Comorbidities, as Defined within Elixhauser Comorbidity Index
Hematologic Cancer vs. Solid Tumor *(Non-Metastatic Only)*
Metastatic Disease
Surgical vs. Non-Surgical Admission
ICU vs. Non-ICU Admission
Length of Stay
Admission via the Emergency Room vs. Other Location
Discharged to Home vs. Other Location
Discharged to Hospice vs. Other Location
Prior Hospitalization
Bone Marrow Transplant

Table 7: Shows 25 potential risk factors identified through the workgroup's review for the *30-Day Unplanned Readmissions for Cancer Patients* measure.

Statistical Methods

The logistic model was fit using SAS/STAT software, Version 9.4 (SAS Institute, Inc. 2017) using the "stepwise" option and maximum likelihood estimation (MLE). Prior to inclusion in the model, we assessed the potential association between the 14 remaining candidate risk adjustors (Table 7, Group E) by calculating the tetrachoric correlation. The potential risk factors were reviewed to identify any variables with correlations of more than 0.5 or less than -0.5, which would require exclusion from the model to avoid multicollinearity (i.e., highly-correlated risk factors).[20] We removed 3 potential model parameters due to high tetrachoric correlations (> 0.5 or < -0.5):

- One Comorbidity: due to high correlation (0.999) with Multiple Comorbidities;
- Admission via the Emergency Room: due to high inverse correlation (-0.6195) with Solid Tumor; and,
- Bone Marrow Transplant: due to high inverse correlation (-0.5027) with Solid Tumor.

We removed 1 potential model parameters due to non-significance:

- Metastatic disease: due to p>0.05.

All logistic model diagnostics, including the c-statistic, Receiver Operating Characteristic (ROC) curve, Hosmer-Lemeshow (H-L) goodness-of-fit test, the likelihood ratio test, and Akaike Information Criterion (AIC), were collected and analyzed prior to selecting the risk adjustment model.[21] Continuous variables, such as "Length of Stay" and "Prior Hospitalization," were analyzed to determine cut points where the rate of readmission increased or decreased. The model was constructed using only binary indicator variables to allow for an intuitive interpretation and application in practice. We calculated odds ratio for each variable (indicating the strength and direction of the association) and used associated confidence intervals to assess significance. An odds ratio > 1.0 or < -1.0, together with a confidence interval excluding 1.00, indicated a significant relationship with the outcome.

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

Below are the potential risk factors evaluated for fit in the risk adjustment model, with observed readmission rates:

Variable Type	Parameter	Readmission Rate		Included in Model?
		Present	Absent	
Factors Included in Model	ICU Stay	20.38%	18.69%	Yes
	Male	19.65%	18.62%	Yes
	Dual-Eligible Status	22.49%	18.32%	Yes
	Surgical Admission	14.91%	21.12%	Yes
	Multiple Comorbidities	20.40%	15.65%	Yes
	Solid Tumor *(Non-Metastatic Only)*	17.52%	20.39%	Yes
	Length of Stay Greater than 3 Days	21.22%	16.07%	Yes
	Age: < 65	*Reference Age*		
	Age: 65-69	18.89%	19.24%	Yes
	Age: 70-74	19.26%	19.14%	Yes
	Age: 75-79	19.28%	19.14%	Yes
	Age: 80-84	18.44%	19.28%	Yes
	Age: 85+	16.30%	19.69%	Yes
	Hospitalization in the Prior 60 Days	26.19%	16.75%	Yes
	Discharged to Home	18.41%	19.89%	Yes
	Discharged to Hospice	2.76%	20.91%	Yes
Factors Excluded from Model	Metastatic Disease	19.65%	18.96%	*No, p>0.05*
	One Comorbidity			*No, high tetrachoric correlation*
	Admission via the Emergency Room			*No, high tetrachoric correlation*
	Bone Marrow Transplant			*No, high tetrachoric correlation*

Table 8: Shows 14 potential risk adjustment variables evaluated for fit in the risk adjustment model for the *30-Day Unplanned Readmissions for Cancer Patients* measure. Data source: Analysis of Medicare SAF (4Q2012-1Q2016), based on data provided by Watson Policy Analysis, 01/13/2017.

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)

Following our conceptual/clinical review of potential risk factors, we further explored the rationale for including 1 SDS factor ("Dual-Eligible Status") in the risk adjustment model. Dual-eligible beneficiaries are, by definition, economically disadvantaged. Thus, "Dual-Eligible Status" can serve as a claims-based proxy for income and other measures of socioeconomic status. Low socioeconomic status is a recognized risk factor for later-stage cancer diagnoses, delayed healthcare receipt, and higher utilization of hospital-based care.[22-24] Moreover, while the evidence is still maturing, socioeconomic status indicators (including private vs. public insurance coverage) have been identified as important predictors of readmissions.[3,10,13]

These findings were further supported through our evaluation of prevalence and fit within the risk adjustment model. "Dual-Eligible Status" was present in 20.17% of index admissions (denominator) and in 23.58% of unplanned readmissions (numerator). The patient-level observed *30-Day Unplanned Readmissions for Cancer Patients* rate was 22.49%, compared with an 18.32% observed rate for all other patients. "Dual-Eligible Status"

was associated with a Chi-Square of 5547.9628 (p<0.001). At the hospital level, the median percentage of dual-eligible patients in the testing dataset was 19.74% (min-max: 0.00%-100.00%; interquartile rate 12.32%-31.25%). Thus, "Dual-Eligible Status" was included in the risk adjustment model.

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*)

The risk adjustment model was tested for adequacy using the dataset described in Section 1.5. Claims were randomly assigned to development and validation samples, using the SAS ranuni function with a seed of "627" to split the data. We used logistic regression analysis to analyze the model's performance, computing c-statistic to evaluate model discrimination and H-L goodness-of-fit test and risk decile plots to evaluate model calibration.[21] Using these statistics, we compared the model performance between the development and validation samples as well as overall adequacy of the risk adjustment model.

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*)**:**

For the development sample (N=1,532,450), the c-statistic is 0.6607 (95% CI: 0.6597, 0.6618). For the validation sample (N=1,535,225), the c-statistic is 0.6609 (95% CI: 0.6588, 0.6630).

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

The H-L Goodness-of-Fit test was X2(df = 8) = 1576.3968 (p<0.001).

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

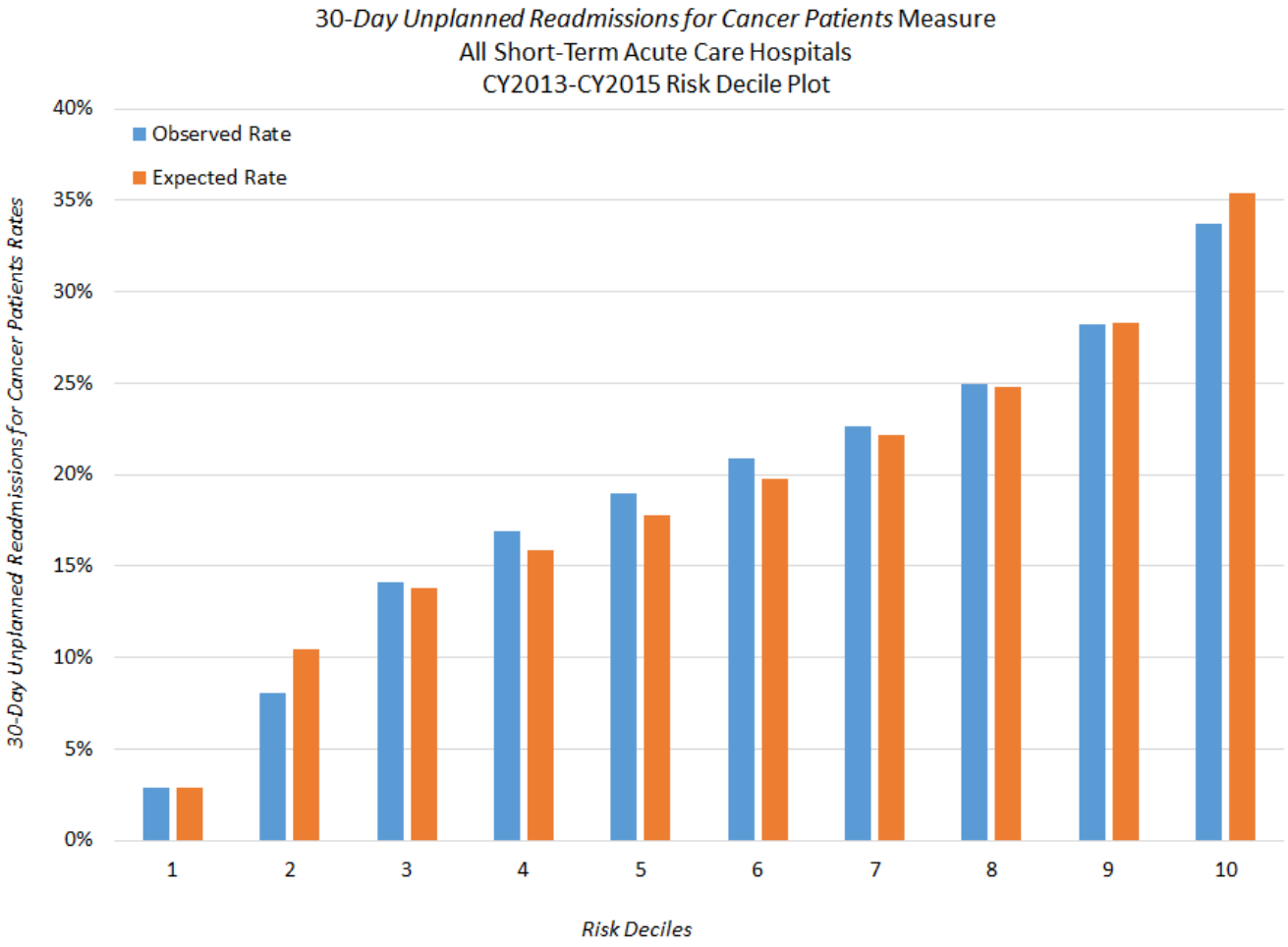


Figure 4: Risk decile plot comparing the observed and expected rates by decile for the risk-adjusted 30-*Day Unplanned Readmissions for Cancer Patients* measure, when applied to 1Q CY2013-4Q CY2015 index admissions for short-term acute care hospitals (i.e., PCHs, short-term acute care PPS hospitals, and CAHs). Data source: Analysis of Medicare SAF (4Q2012-1Q2016), based on data provided by Watson Policy Analysis, 01/13/2017.

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*)

The c-statistic measures how well the model discriminates between patients with and without the outcome, when compared with random assignment. A c-statistic of 0.5 suggests that the model has poor predictive power, while a c-statistic of 1.0 implies that the outcome is solely related to patient-level factors. The c-statistic here is 0.6607 (95% CI: 0.6597, 0.6618), indicating fair discrimination for the development and validation models. Likewise, the wide range between the lowest and highest deciles indicates that the model discriminates between high- and low-risk patients.

The H-L Goodness-of-Fit test yielded a significant value ($p<0.001$), which indicates potential fit issues. This is not uncommon with models that are overpowered due to large datasets, as is the case here. A significant value for the H-L test suggests that we reject the assumption of perfect fit between the models. However, with large datasets, the H-L statistic can magnify relatively small differences between observed and expected rates and

imply a statistically significant degree of miscalibration. When viewed within the context of the c-statistic and the risk decile plots, the significant H-L statistic does not suggest that the model has poor calibration.[25,26]

The risk decile plots demonstrate that the model performs adequately, with similar observed and predicted values in each decile. Together, the discrimination and calibration tests confirm the adequacy of the risk adjustment model in controlling for differences in patient-level risk factors.

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*)

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)*

We applied the logistic regression model to the testing dataset as an assessment of statistical significance. The final unadjusted and risk-adjusted rates were calculated across all hospitals in the testing dataset and separately for 2015. To calculate risk-adjusted rates, we used the equation described in Section 2b4.1.1. First, we calculated observed (unadjusted) rates by hospital. We used the logistic regression model described in Section 2b4 to calculate the expected rate for each hospital. Then, we calculated observed/expected ratios for each hospital and multiplied those rates by the national observed rate to yield risk-adjusted rates by hospital. We calculated the confidence interval (CI) for each hospital's score to interpret each hospital's performance, when compared with the national observed rate. Performance scores were interpreted as follows:

- If the confidence interval contained the national average, the hospital's performance rate was interpreted as *no better or worse* than the national average;
- If the confidence interval was *lower* than the national average, the hospital's performance rate was interpreted as *better* than the national average; and,
- If the confidence interval was *higher* than the national average, the hospital's performance rate was interpreted as *worse* than the national average.

We also generated histograms to visualize the distribution of unadjusted and risk-adjusted scores for CY2013-CY2015.

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*)

	All Short-Term Acute Care Hospitals	PPS-Exempt Cancer Hospitals	Short-Term Acute Care PPS Hospitals	Critical Access Hospitals
Number of Hospitals	4,974	11	3,617	1,346
Number of Admissions (Denominator)	3,067,675	73,159	2,934,917	59,599
Number of Unplanned Readmissions (Numerator)	587,915	15,724	563,701	8,490
National Rates				
Unadjusted 30-Day Unplanned Readmission Rate	19.16%	21.49%	19.21%	14.25%
Risk-Adjusted 30-Day Unplanned Readmission Rate	19.13%	20.20%	19.20%	14.13%
Hospital-Level Rates				
Mean (Standard Deviation)	16.61% (8.07%)	18.79% (3.06%)	17.78% (6.98%)	13.44% (9.79%)
Range (Min-Max)	0.00%-100.00%	12.22%-21.57%	0.00%-100.00%	0.00%-99.62%
Quartile Range	7.35%	2.97%	5.38%	11.39%
Minimum	0.00%	12.22%	0.00%	0.00%
25th percentile	13.21%	18.01%	15.44%	7.20%
50th percentile	17.68%	20.83%	18.45%	13.03%
75th percentile	20.56%	20.98%	20.82%	18.60%
Maximum	100.00%	21.57%	100.00%	99.62%

Table 9: Shows the results of the logistic regression model for the 30-*Day Unplanned Readmissions for Cancer Patients* measure, when applied to 1Q CY2013-4Q CY2015 index admissions for short-term acute care hospitals (i.e., PCHs, short-term acute care PPS hospitals, and CAHs). Data source: Analysis of Medicare SAF (4Q2012-1Q2016), based on data provided by Watson Policy Analysis, 01/13/2017.

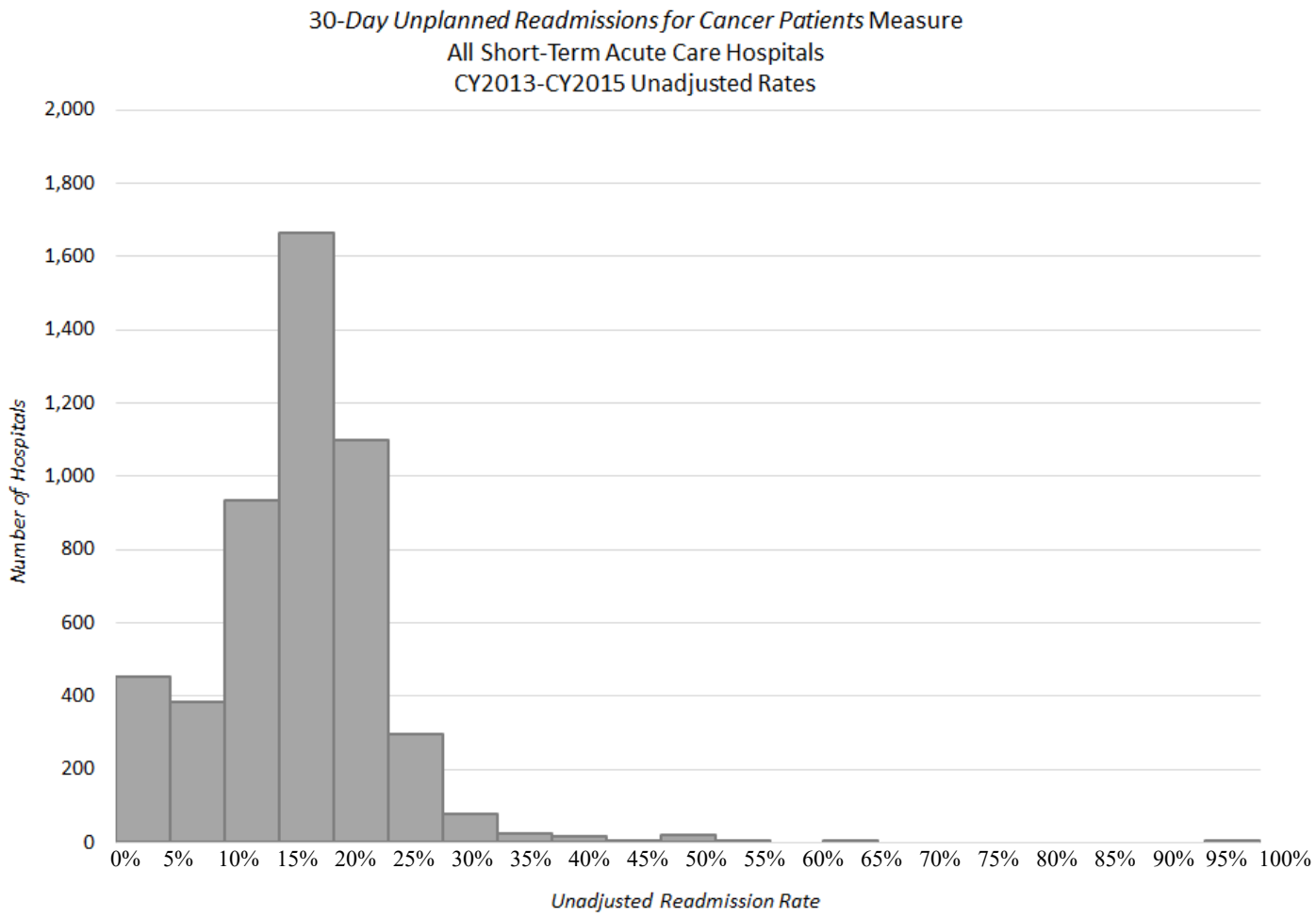


Figure 5: Histogram showing the distribution of unadjusted rates by hospital for the 30-*Day Unplanned Readmissions for Cancer Patients* measure, when applied to 1Q CY2013-4Q CY2015 index admissions for short-term acute care hospitals (i.e., PCHs, short-term acute care PPS hospitals, and CAHs). Data source: Analysis of Medicare SAF (4Q2012-1Q2016), based on data provided by Watson Policy Analysis, 01/13/2017.

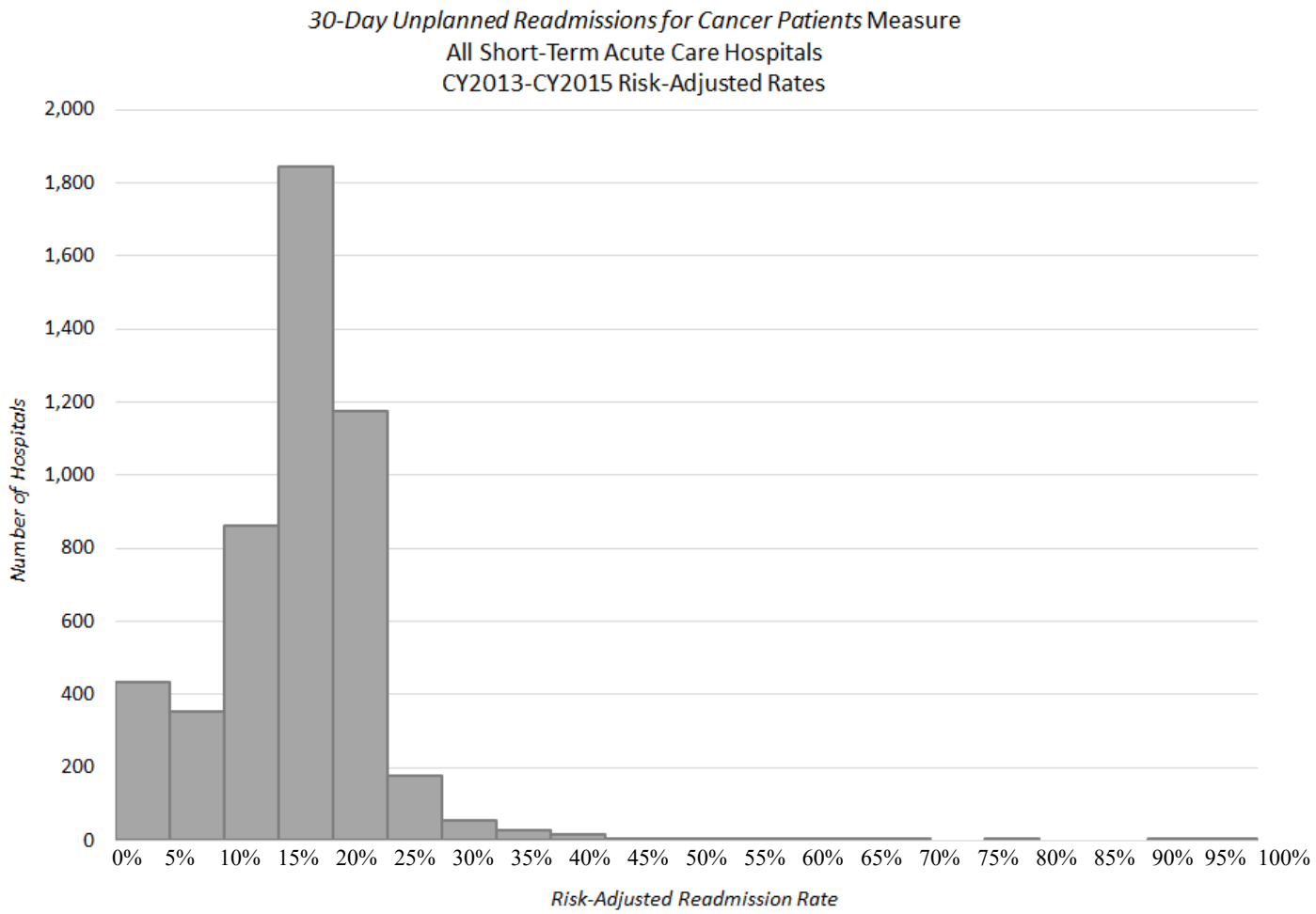


Figure 6: Histogram showing the distribution of risk-adjusted rates by hospital for the 30-*Day Unplanned Readmissions for Cancer Patients* measure, when applied to 1Q CY2013-4Q CY2015 index admissions for short-term acute care hospitals (i.e., PCHs, short-term acute care PPS hospitals, and CAHs). Data source: Analysis of Medicare SAF (4Q2012-1Q2016), based on data provided by Watson Policy Analysis, 01/13/2017.

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 30-*Day Unplanned Readmissions for Cancer Patients* measure was able to detect hospitals with better and worse than the national average performance rate. For CY2013-CY2015, the unadjusted readmission rates ranged from 0.00%-100%, with a median rate of 17.32%. Half of the hospitals fell within the interquartile range of 12.50%-20.80%. The mean unadjusted rate was 16.54% (SD=8.24%). The risk-adjusted rates had the same overall range and a narrower interquartile range (13.21%-20.56%). The mean risk-adjusted rate was 16.61% (SD=8.07%). The histograms for unadjusted and risk-adjusted rates showed performance rates skewed right and few high outliers, consistent with the statistics described above. Likewise, when analyzed in total and for CY2015 individually, we observed that over half of all index claims had performance *no better or worse* than the national average. This conforms with the narrow interquartile range we observed. Together, these statistics indicate that there are opportunities to utilize this measure to reduced unplanned readmissions in cancer patients, making it useful for performance improvement and public reporting.

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

References

1. Hu J, Gonsahn MD, Nerenz DR. Socioeconomic status and readmissions: evidence from an urban teaching hospital. *Health Aff (Millwood).* 2014;33(5):778-785.
2. Joynt KE, Orav EJ, Jha AK. Thirty-day readmission rates for Medicare beneficiaries by race and site of care. *JAMA.* 2011;305(7):675-681.
3. Weissman JS, Stern RS, Epstein AM. The impact of patient socioeconomic status and other social factors on readmission: a prospective study in four Massachusetts hospitals. *Inquiry.* 1994;31(2):163-172.
4. Jencks SF, Williams MV, Coleman EA. Rehospitalizations among patients in the Medicare fee-for-service program. *N Engl J Med.* 2009;360(14):1418-1428.
5. Bernheim SM, Parzynski CS, Horwitz L, et al. Accounting For Patients' Socioeconomic Status Does Not Change Hospital Readmission Rates. *Health Aff (Millwood).* 2016;35(8):1461-1470.
6. Manzano JG, Luo R, Elting LS, George M, Suarez-Almazor ME. Patterns and predictors of unplanned hospitalization in a population-based cohort of elderly patients with GI cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology.* 2014;32(31):3527-3533.
7. Shah SP, Xu T, Hooker CM, et al. Why are patients being readmitted after surgery for esophageal cancer? *J Thorac Cardiovasc Surg.* 2015;149(5):1384-1389; discussion 1389-1391.
8. Fernandez FG, Khullar O, Force SD, et al. Hospital readmission is associated with poor survival after esophagectomy for esophageal cancer. *Ann Thorac Surg.* 2015;99(1):292-297.
9. Valero-Elizondo J, Kim Y, Prescott JD, et al. Incidence and Risk Factors Associated with Readmission After Surgical Treatment for Adrenocortical Carcinoma. *J Gastrointest Surg.* 2015;19(12):2154-2161.
10. Wilbur MB, Mannschreck DB, Angarita AM, et al. Unplanned 30-day hospital readmission as a quality measure in gynecologic oncology. *Gynecol Oncol.* 2016;143(3):604-610.
11. Dickinson H, Carico C, Nuno M, et al. Unplanned readmissions and survival following brain tumor surgery. *J Neurosurg.* 2015;122(1):61-68.
12. Uppal S, Penn C, Del Carmen MG, Rauh-Hain JA, Reynolds RK, Rice LW. Readmissions after major gynecologic oncology surgery. *Gynecol Oncol.* 2016;141(2):287-292.
13. Nakayama JM, Ou JP, Friedman C, Smolkin ME, Duska LR. The Risk Factors of Readmission in Postoperative Gynecologic Oncology Patients at a Single Institution. *Int J Gynecol Cancer.* 2015;25(9):1697-1703.
14. Rousson V, Gasser T, Seifert B. Assessing intrarater, interrater and test-retest reliability of continuous measurements. *Stat Med.* 2002;21(22):3431-3446.
15. Bartko JJ. The intraclass correlation coefficient as a measure of reliability. *Psychol Rep.* 1966;19(1):3-11.
16. Evans JD. *Straightforward statistics for the behavioral sciences.* Pacific Grove: Brooks/Cole Pub. Co.; 1996.
17. Cohen J. A power primer. *Psychol Bull.* 1992;112(1):155-159.
18. Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment.* 1994;6(4):284-290.
19. Agency for Healthcare Research and Quality. Elixhauser Comorbidity Software, Version 3.7. Available at: https://www.hcup-us.ahrq.gov/toolssoftware/comorbidity/comorbidity.jsp. Accessed January 13, 2017.
20. van Walraven C, Austin PC, Jennings A, Quan H, Forster AJ. A modification of the Elixhauser comorbidity measures into a point system for hospital death using administrative data. *Med Care.* 2009;47(6):626-633.
21. Bewick V, Cheek L, Ball J. Statistics review 14: Logistic regression. *Crit Care.* 2005;9(1):112-118.
22. Mitchell AD, Inglis KM, Murdoch JM, Porter GA. Emergency room presentation of colorectal cancer: a consecutive cohort study. *Ann Surg Oncol.* 2007;14(3):1099-1104.
23. Hargarten SW, Richards MJ, Anderson AJ. Cancer presentation in the emergency department: a failure of primary care. *Am J Emerg Med.* 1992;10(4):290-293.
24. Amri R, Bordeianou L, Sylla P, Berger DL. Colon Cancer Surgery Following Emergency Presentation: Effects on Admission and Stage-adjusted Outcomes. *The American Journal of Surgery.* 2014.
25. Kramer AA, Zimmerman JE. Assessing the calibration of mortality benchmarks in critical care: The Hosmer-Lemeshow test revisited. *Crit Care Med.* 2007;35(9):2052-2056.
26. Feudtner C, Hexem KR, Shabbout M, Feinstein JA, Sochalski J, Silber JH. Prediction of pediatric death in the year after hospitalization: a population-level retrospective cohort study. *J Palliat Med.* 2009;12(2):160-169.

3. Feasibility

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

3a. Byproduct of Care Processes

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

3a.1. **Data Elements Generated as Byproduct of Care Processes.**
Generated or collected by and used by healthcare personnel during the provision of care (e.g., blood pressure, lab value, diagnosis, depression score), 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*) Update this field for **maintenance of endorsement**.
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.** For **maintenance of endorsement**, if this measure is not an eMeasure (eCQM), please describe any efforts to develop an eMeasure (eCQM).

3b.3. **If this is an eMeasure, provide a summary of the feasibility assessment in an attached file or make available at a measure-specific URL. Please also complete and attach the NQF Feasibility Score Card.**
Attachment:

3c. Data Collection Strategy

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

3c.1. **Required for maintenance of endorsement. Describe difficulties (as a result of testing and/or operational use of the measure) regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.**
IF a PRO-PM, consider implications for both individuals providing PRO data (patients, service recipients, respondents) and those whose performance is being measured.
N/A

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)*.
N/A

4. Usability and Use

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

4a. Accountability and Transparency

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

4.1. Current and Planned Use

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

Specific Plan for Use	Current Use (for current use provide URL)
	Public Reporting 2) Annual Hospital Ratings for Colon and Lung Cancer Surgery-US News&World Report http://health.usnews.com/health-news/blogs/second-opinion/articles/2016-07-07/methodology-updated-for-ratings-in-procedures-and-conditions Payment Program Accountable Care Program-Moffitt Cancer Center/Florida Blue https://www.moffitt.org/ Quality Improvement (external benchmarking to organizations) Vizient (neé University HealthSystem Consortium, or UHC) Clinical Data Base/Resource Manager https://www.vizientinc.com/ Quality Improvement (Internal to the specific organization) City of Hope Comprehensive Cancer Center http://www.cityofhope.org/homepage University of Miami Sylvester Comprehensive Cancer Center http://sylvester.org/ Seattle Cancer Care Alliance http://www.seattlecca.org/

4a.1. For each CURRENT use, checked above (update for maintenance of endorsement), provide:

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

1) Accountable Care Program: Moffitt Cancer Center
Purpose: Moffitt has incorporated the 30-day Unplanned Readmissions for Cancer Patients measure in the first-ever cancer-specific accountable care program with Florida Blue. Moffitt is using this measure to identify opportunities to reduce unplanned readmissions for Florida Blue beneficiaries as part of broader efforts to improve individual patient care and decrease costs of care.
Geographic area: Florida.
Level of measurement and setting: Facility/hospital.

2) US News&World Report (USNWR): Annual Hospital Ratings for Lung and Colon Cancer Surgery
Purpose: USNWR adopted the 30-day Unplanned Readmissions for Cancer Patients measure for use in its annual hospital ratings for colon and lung cancer surgeries. The measure was empirically selected after reviewing as many as three candidate readmission measures for the cohort, and with the recommendation of a volunteer medical advisory panel convened to advise USNWR on approaches to evaluating cancer care.
Geographic area: This measure is applied to all hospitals included in USNWR's annual hospital ratings for colon and lung cancer surgeries.
Level of measurement and setting: Facility/hospital.

3) Vizient: Quality Improvement with Benchmarking
Purpose: PCHs and other comprehensive cancer centers actively use this measure to compare their performance against other

members' performance for purposes of benchmarking and identification of internal performance improvement opportunities.
Geographic area: This measure is available for use by Vizient members throughout the United States that submit data to the CDB/RM.
Level of measurement and setting: Facility/hospital, with stratification and drill-down capability for the reporting facility

4a) Quality Improvement: City of Hope Comprehensive Cancer Center
Purpose: City of Hope uses the measure in monthly quality improvement reports for hospital leadership.
Geographic area: Southern California (Los Angeles area).
Level of measurement and setting: Facility/hospital.

4b) Quality Improvement: University of Miami Sylvester Comprehensive Cancer Center
Purpose: Sylvester uses the measure to help guide care decisions in discharge planning.
Geographic area: Southern Florida (Miami area).
Level of measurement and setting: Facility/hospital.

4c) Quality Improvement: Seattle Cancer Care Alliance
Purpose: Seattle Cancer Care Alliance uses the measure as a comparison for the HWR measure (NQF #1789) to demonstrate sensitivity of treating cancer patients as a separate category for systems reporting.
Geographic area: Pacific Northwest (Seattle area).
Level of measurement and setting: Facility/hospital.

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?*)

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.*)
This measure was included in CMS' 2014 Measures Under Consideration (MUC) list and received conditional support from the Measure Applications Partnership (MAP) Hospital Work Group, pending NQF endorsement. It is our expectation this measure will be included in future rulemaking, potentially as early as the FY 2018 IPPS Proposed Rule.

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. Refer to data provided in 1b but do not repeat here. Discuss any progress on improvement (trends in performance results, number and percentage of people receiving high-quality healthcare; Geographic area and number and percentage of accountable entities and patients included.)
If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

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. **Please explain any unexpected findings (positive or negative) during implementation of this measure including unintended impacts on patients.**
No unintended negative consequences were identified during testing. This is a passive surveillance approach with no attached intervention.

4c.2. **Please explain any unexpected benefits from implementation of this measure.**
The measure can serve as an impetus for quality improvement in discharge planning for cancer patients.

4d1.1. **Describe how performance results, data, and assistance with interpretation have been provided to those being measured or other users during development or implementation.**
How many and which types of measured entities and/or others were included? If only a sample of measured entities were included, describe the full population and how the sample was selected.
An earlier version of this measure, which examines unplanned readmissions to the discharging facility only, is readily available to any Vizient member. Many quality officers at PCHs institutions routinely access the data for purposes of internal quality reporting. With the revised measure specifications, it is anticipated that public reporting through the PCHQR will allow for greater access to performance data. Moreover, we believe that the measure has broad applicability to cancer patients treated in other short-term acute care hospitals and can, therefore, be adopted for other public reporting programs.

4d1.2. **Describe the process(es) involved, including when/how often results were provided, what data were provided, what educational/explanatory efforts were made, etc.**
This measure was developed principally for the PCHQR and has not yet been adopted for the program. Additional information is forthcoming following its adoption for public reporting.

4d2.1. **Summarize the feedback on measure performance and implementation from the measured entities and others described in 4d.1.**
Describe how feedback was obtained.
Please see comments above.

4d2.2. **Summarize the feedback obtained from those being measured.**
Please see comments above.

4d2.3. **Summarize the feedback obtained from other users**
Please see comments above.

4d.3. **Describe how the feedback described in 4d.2 has been considered when developing or revising the measure specifications or implementation, including whether the measure was modified and why or why not.**
Please see comments above.

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)**
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 of Related Measures
The measure specifications are harmonized with related measures;

OR
The differences in specifications are justified

5a.1. **If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s):**
Are the measure specifications harmonized to the extent possible?
No

5a.2. **If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.**
The 30-Day Unplanned Readmissions for Cancer Patients measure has a different target population from the HWR measure (NQF #1789), which expressly excludes admissions to PCHs, noting that the PCHs care for a unique patient population that is challenging to compare to other hospitals. Moreover, the HWR measure excludes non-surgical admissions for cancer patients because the outcomes do not correlate well with outcomes for other admissions. Due to the different target populations for each measure, it does not require harmonization with the HWR measure (NQF #1789).

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.)

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:** 2016_12_22_UnplannedReadm_Cancer_Appendix.pdf

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** Seattle Cancer Care Alliance
Co.2 **Point of Contact:** Barb, Jagels, bjagels@seattlecca.org, 206-288-2127-
Co.3 **Measure Developer if different from Measure Steward:** Alliance of Dedicated Cancer Centers
Co.4 **Point of Contact:** Tracy, Spinks, tespinks@mdanderson.org, 713-563-2198-

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.
Joseph M. Flynn, MD Arthur G. James Cancer Hospital and Richard J. Solove Research Institute
Kristen Johnson, MHA Arthur G. James Cancer Hospital and Richard J. Solove Research Institute
Linda Lane, RHIA, CPHQ Arthur G. James Cancer Hospital and Richard J. Solove Research Institute
Tonja Plew, BSN, RN Arthur G. James Cancer Hospital and Richard J. Solove Research Institute
Julie Rader, BSN, RN Arthur G. James Cancer Hospital and Richard J. Solove Research Institute
Carl Schmidt, MD Arthur G. James Cancer Hospital and Richard J. Solove Research Institute
Susan White, PhD, RHIA, CHDA Arthur G. James Cancer Hospital and Richard J. Solove Research Institute (neé Health Policy Analytics, LLC)
Angie Wolf-Erdlitz, BS, MGS, RN Arthur G. James Cancer Hospital and Richard J. Solove Research Institute
Denise Morse, MBA City of Hope Comprehensive Cancer Center

Laura Crocitto, MD, MHA City of Hope Comprehensive Cancer Center
Merle Smith, RN, BSBA, MSN City of Hope Comprehensive Cancer Center
Steve Flaherty, MPH Dana-Farber Cancer Institute
Apar Gupta, MD Dana-Farber Cancer Institute
Audrey Holland, PhD Duke Cancer Institute
Steve Power, MBA, CSQE, CQA Duke Cancer Institute

Joyce M. Kane, MSN, RN, CPHQ, RHIT, CTR Johns Hopkins Sidney Kimmel Cancer Center
Sarah Berger, MBA Memorial Sloan Kettering Cancer Center
Steve Martin, MD Memorial Sloan Kettering Cancer Center
Dhruvkumar Patel, MS Memorial Sloan Kettering Cancer Center
Kathleen Trainor Memorial Sloan Kettering Cancer Center
Jennifer Snide, MS Norris Cotton Cancer Center at Dartmouth-Hitchcock
Keith Eaton, MD, PhD Seattle Cancer Care Alliance
Lois Helbert, RN Seattle Cancer Care Alliance
Paul Hendrie, MD, PhD (Chair) Seattle Cancer Care Alliance
Barb Jagels, RN, MHA, CPHQ Seattle Cancer Care Alliance
Tracy Kusnir-Wong, MBA Seattle Cancer Care Alliance
Paul Reitz Seattle Cancer Care Alliance
Gloria (Gigi) Campos, MSIE University of Miami Sylvester Comprehensive Cancer Center
Laurian Walters, BS University of Miami Sylvester Comprehensive Cancer Center
Doug Browning, MBA The University of Texas MD Anderson Cancer Center
Mercy Cherian, RN, MSN, OCN, RN-BC The University of Texas MD Anderson Cancer Center
Sobha George, MSN, CEN, RN The University of Texas MD Anderson Cancer Center
Tiffany Jones, DrPh The University of Texas MD Anderson Cancer Center
Bettina Patterson, MSN, RN, OCN The University of Texas MD Anderson Cancer Center
Shalonda Smith, RN, BSN The University of Texas MD Anderson Cancer Center
Tracy Spinks, BBA The University of Texas MD Anderson Cancer Center
Colleen Tallant, MS The University of Texas MD Anderson Cancer Center
Ron Walters, MD, MBA, MS, MHA The University of Texas MD Anderson Cancer Center
Lavinia Zanaj, MHA, RN The University of Texas MD Anderson Cancer Center
Afsaneh Barzi, MD, PhD USC Norris Comprehensive Cancer Center
Stephanie Buia Amport, MBA, CPHQ Yale Cancer Center
Lisa Truini-Pittman, RN, MPH Yale Cancer Center

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:** 12, 2016
Ad.4 **What is your frequency for review/update of this measure?** Annual
Ad.5 **When is the next scheduled review/update for this measure?** 12, 2017

Ad.6 **Copyright statement:** N/A
Ad.7 **Disclaimers:** N/A

Ad.8 **Additional Information/Comments:** In January 2016, the ADCC submitted this measure for consideration by the NQF All-Cause Admissions and Readmissions Project 2015-2017 TEP. The TEP was convened June 8-9, 2016 to review all submitted measures and provide recommendations regarding measure endorsement. During the review, the TEP expressed enthusiasm for a cancer-specific readmissions measure but did not support endorsement of the measure, as submitted. The TEP noted concerns related to the limited testing population and the measure's focus on unplanned readmissions to the discharging hospital only.

Following the recommendation of the TEP, the ADCC broadened the measure to capture readmissions of cancer patients from and to any short-term acute care hospital (PCHs, short-term acute care PPS hospitals, and CAHs) and pursued additional testing of the measure using Medicare claims data. This expansion produced unplanned readmissions rates of patients discharged from PCHs and readmitted to any short-term acute care hospital (PCH, short-term acute care PPS hospital, or CAH). Additionally, it provided comparative rates of unplanned readmissions of cancer patients for non-PCH short-term acute care hospitals (i.e., short-term acute care PPS hospitals and CAHs). We believe that the measure has broad applicability to cancer patients treated in other short-term acute care hospitals and can be successfully adopted for the PCHQR and other public reporting programs.

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)

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)

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

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

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 <u>all</u> 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 <u>measured entities</u> 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.

For reliability testing (Section 2a2)
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**.**

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

For validiation of the measure's risk model (Section 2b2):
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			

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)
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)

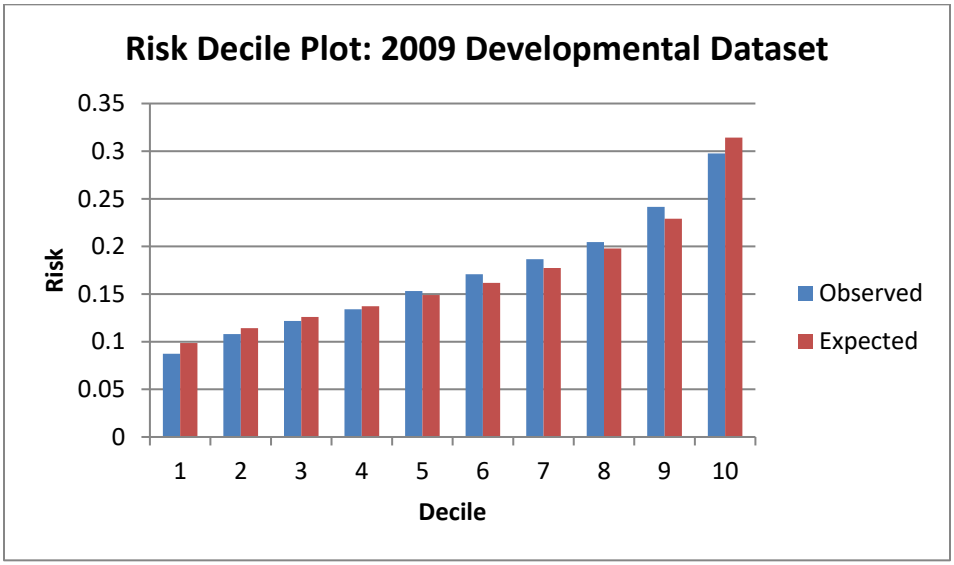
2010 validation cohort:
C-statistic = 0.63
Predictive ability (lowest decile %, highest decile %): (8.4, 30.3)

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